

AD-A171 935

WAVELENGTH INDEPENDENT OPTICAL LITHOGRAPHY(U) CORNELL
UNIV ITHACA NY SCHOOL OF APPLIED AND ENGINEERING
PHYSICS A LEWIS 06 JUN 86 AFOSR-TR-86-0358

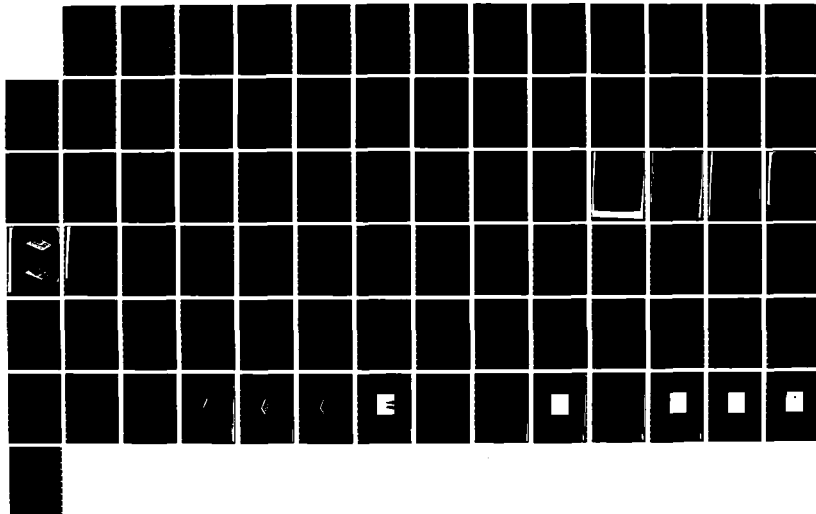
1/1

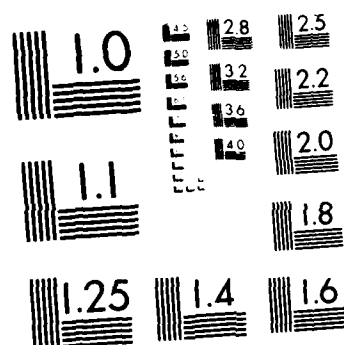
UNCLASSIFIED

AFOSR-84-0314

F/G 20/6

NL





MICROCOPY RESOLUTION TEST CHART

2

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
AFOSR-TR. 86-0558		
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED
"WAVELENGTH INDEPENDENT OPTICAL LITHOGRAPHY"		ANNUAL FINAL, 01SEP84 - 31AUG85
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s)
Aaron Lewis		AFOSR-84-0314
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
CORNELL UNIVERSITY School of Applied and Engineering Physics Ithaca NY 14853		61102F 2306/B2
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE
		06 June 86
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
AFOSR/NE Bolling AFB DC 20332-6448		Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)		
DISTRIBUTION STATEMENT A Approved for public release Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
Approved for public release, distribution unlimited		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
ABSTRACT		
<p>The transmission of light through an infinite slit in a thick, perfectly conducting screen is investigated. The spatial distribution of the near-field energy flux is determined through the formulation of four coupled integral equations, which are solved numerically. Transmission coefficients calculated by this method are in agreement with those determined by an alternative formulation. The results theoretically demonstrate the feasibility of near-field super-resolution microscopy, in which the collimated radiation passed by an aperture is used to circumvent the diffraction limit of conventional optics, and further suggest the feasibility of near-field super-resolution acoustic imaging.</p>		

AD-A171 935

DTIC FILE COPY

DTIC
ELECTE
SEP 15 1986
S B

DISCLAIMER NOTICE

**THIS DOCUMENT IS BEST QUALITY
PRACTICABLE. THE COPY FURNISHED
TO DTIC CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO NOT
REPRODUCE LEGIBLY.**

AFOSR-TR. 86-0558

ANNUAL TECHNICAL REPORT

AFOSR GRANT # 84-0314

WAVELENGTH INDEPENDENT OPTICAL LITHOGRAPHY

Approved for public release,
distribution unlimited

AARON LEWIS

SCHOOL OF APPLIED AND ENGINEERING PHYSICS

CORNELL UNIVERSITY

ITHACA, NEW YORK 14853

This report will be broken into two parts. The first part will deal with the theory underlying the principles of wavelength independent optical lithography and microscopy and the second part will deal with the experimental realization of these theoretical principles.



Acc	✓
AGL	
Dist	
A-1	

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DTIC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12.
Distribution is unlimited.
MATTHEW J. KENNER
Chief, Technical Information Division

PART I

APPLIED OPTICS (In press)

Near-Field Diffraction by a Slit: Implications for Super-Resolution
Microscopy

E. Betzig, A. Harootunian, A. Lewis, and M. Isaacson

Cornell University -
School of Applied and Engineering Physics
Ithaca, NY 14853

ABSTRACT

The transmission of light through an infinite slit in a thick, perfectly conducting screen is investigated. The spatial distribution of the near-field energy flux is determined through the formulation of four coupled integral equations, which are solved numerically. Transmission coefficients calculated by this method are in agreement with those determined by an alternative formulation. The results theoretically demonstrate the feasibility of near-field super-resolution microscopy, in which the collimated radiation passed by an aperture is used to circumvent the diffraction limit of conventional optics, and further suggest the feasibility of near-field super-resolution acoustic imaging.

INTRODUCTION

The transmission of radiation through small apertures has been investigated for a number of years. Bethe¹ treated the case of diffraction by a round hole in an infinitely thin, perfectly conducting screen in the limit where the hole was small compared to the wavelength, and considered the applicability of his solution to frequency behavior and power flow in microwave cavities. More recently, research in aperture diffraction has been aimed at obtaining a better understanding of the effects of nuclear-induced electromagnetic pulses.² However, these studies have not dealt with the near-field spatial distribution of radiation transmitted past the aperture. The development by Ash and Nicholls³ of a scanning microwave microscope with $\lambda/60$ resolution provides the incentive to apply aperture diffraction theory to a new problem: the determination of the practical limits of near-field super-resolution microscopy. Although in this paper we will consider the

specific case of optical microscopy, the results are applicable to other regimes of the electromagnetic spectrum.

In the near-field microscope, an illuminated subwavelength aperture is used as a light source. This aperture is then placed within the near field relative to an object and an image is generated by scanning the aperture relative to the object (see Fig. 1). From a theoretical standpoint, the most interesting questions concern the degree of collimation of the radiation immediately past the screen and the rate of spreading of the radiation farther from the screen, which would determine the resolution capabilities and positioning requirements of a super-resolution instrument. To address these questions, a two-dimensional aperture model will be considered, i.e., an infinite slit in a thick screen of perfect conductivity.

The infinite slit geometry has already been considered by a number of researchers. Harrington and Auckland² have analyzed the problem using the method of moments and have developed a simple equivalent circuit in the narrow slit limit (width $\ll \lambda$). However, their results are concerned primarily with transmission coefficients rather than the spatial distribution of the radiation. Neerhoff and Mur³ have used a Green's function formulation for the problem. Indeed, their derivation will be used for the analysis here, and is summarized below. The difference between this work and that of Neerhoff and Mur lies in the manner in which the model is applied. First, the smallest slits considered by Neerhoff and Mur were roughly $\lambda/2$ in width, whereas much smaller slits (width $=\lambda/10$) are of interest in super-resolution applications. Second, although the authors compare the electromagnetic field amplitudes in the plane of the slit with similar amplitudes farther from the screen, the down-screen distances chosen for comparison are generally beyond the range of collimation as determined by the more thorough calculations included in this

work. Finally, when considering illumination for microscopic or photolithographic applications, the energy flux transmitted by the slit is of greater physical significance than are the field amplitudes calculated by Neerhoff and Mur.

In short, the theoretical development of Neerhoff and Mur will be used to find the fields at the boundaries of a number of narrow slits and these fields will then be used to calculate the spatial distribution of the transmitted energy flux within the near field. To check the consistency of the results, transmission coefficients will be calculated and compared to the findings of Harrington and Auckland. The implications of the results for super-resolution microscopy will then be discussed.

THEORETICAL DEVELOPMENT

Initial Formulation

The following formulation summarizes the work of Neerhoff and Mur. The notation used has also been largely adopted from that work.

The configuration appropriate for the formulation of the problem is shown in Fig. 2. The slit, with a width $2a$, is chosen parallel to the y axis and is contained in a screen of thickness b . The space surrounding the slit is divided into three regions. Region I ($|x| \leq a$, $|y| \leq a$, $b < z \leq \infty$) contains the space above the slit and includes the incident electromagnetic wave. Region II ($|x| < a$, $|y| \leq a$, $0 < z < b$) contains the space inside the slit. Finally, region III ($|x| \leq a$, $|y| \leq a$, $-\infty < z < 0$) contains the space below the slit and includes the transmitted wave. Different dielectrics can exist in the different regions.

The incident plane wave in region I propagates in the xz plane at an angle θ with respect to the z axis. The magnetic field is assumed to be time harmonic and both polarized and constant in the y direction:

$$\vec{H}(x, y, z, t) = U(x, z) \exp(-i\omega t) \hat{e}_y \quad (1)$$

These restrictions are placed for two reasons. First, elementary waveguide calculations indicate that at least one propagating mode exists for this polarization, no matter how narrow the slit. Hence, the arrangement is of physical interest since large transmission coefficients can be anticipated. Second, Maxwell's equations can be used to find the electric field from $U(x,z)$:

$$E_x(x,z) = -\frac{1}{\omega\epsilon} \partial_z U(x,z) \quad (2a)$$

$$E_y(x,z) = 0 \quad (2b)$$

$$E_z(x,z) = \frac{1}{\omega\epsilon} \partial_x U(x,z) \quad (2c)$$

Thus, the restrictions in Eq. (1) reduce the diffraction problem to one involving a single scalar field in only two dimensions.

The field is represented by $U_j(x,z)$ ($j=1,2,3$) in each of the three regions. The time harmonic behavior indicates that the field satisfies the Helmholtz equation:

$$(\nabla^2 + k_j^2)U_j = 0 \quad (j=1,2,3) \quad (3)$$

In region I the field is further decomposed into three components:

$$U_1(x,z) = U^i(x,z) + U^r(x,z) + U^d(x,z) \quad (5)$$

each of which satisfy:

$$(\nabla^2 + k_1^2)U_j = 0 \quad (j=i,r,d) \quad (6)$$

U^i represents the incident field, which, for simplicity, is assumed to be a plane wave of unit amplitude:

$$U^i(x,z) = e^{ik_1(x \sin \theta - z \cos \theta)} \quad | \theta | < \pi/2 \quad (7)$$

U^r denotes the field that would be reflected if there were no slit in the screen, and thus satisfies:

$$U^r(x, z) = U^l(x, 2b - z) \quad (9)$$

U^d describes the diffracted field in region I due to the presence of the slit.

For a perfectly conducting screen, the tangential \vec{E} vanishes at the surface, so that:

$$\partial_n U = 0 \text{ at the conductor surface} \quad (9)$$

where n is the normal vector at the surface. Further boundary conditions result from assuming that the diffracted fields vanish at infinity:

$$U_d, U_3 \rightarrow o\left(\frac{1}{R}\right) \text{ as } R \rightarrow \infty \quad (10a)$$

$$\partial_n U_d, \partial_n U_3 \rightarrow o\left(\frac{1}{R}\right) \text{ as } R \rightarrow \infty \quad (10b)$$

The final boundary conditions are valid at the upper and lower slit boundaries:

$$U_1(x, z)|_{z=b^+} = U_2(x, z)|_{z=b^-} \quad \text{for } |x| < a \quad (11a)$$

$$U_2(x, z)|_{z=0^+} = U_1(x, z)|_{z=0^-} \quad \text{for } |x| < a \quad \text{continuity of tangential } H \quad (11b)$$

$$\epsilon_2 \partial_z U_1(x, z)|_{z=b^+} = \epsilon_1 \partial_z U_2(x, z)|_{z=b^-} \quad \text{for } |x| < a \quad (12a)$$

$$\epsilon_1 \partial_z U_2(x, z)|_{z=0^+} = \epsilon_2 \partial_z U_1(x, z)|_{z=0^-} \quad \text{for } |x| < a \quad \text{continuity of tangential } E \quad (12b)$$

With the above set of equations and boundary conditions, a unique solution exists for the diffraction problem.

Development of Integral Field Representations

To find the field, the two dimensional Green's theorem is applied, with one function given by $U(x, z)$ and the other by a standard Green's function:

$$(\nabla^2 + k_j^2) G_j = -\delta(x - x', z - z') \quad (j=1, 2, 3) \quad (13)$$

where (x, z) refers to a field point of interest and x', z' are integration variables. Since G_j satisfies Eq. (3), Green's theorem reduces to:

$$U(x, z) = \int_{\text{Boundary}} (G \partial_n U - U \partial_n G) dS \quad (14)$$

where the boundaries are shown in Fig. 3.

By virtue of Eq. (9), more is known about $\partial_n U$ at the boundaries than is known about U itself. Hence, we impose boundary conditions on G :

$$\partial_z G_1(x, z) \big|_{z=b} = 0 \quad \text{for } |x| < \infty \quad (15a)$$

$$\partial_z G_2(x, z) \big|_{z=0} = 0 \quad \text{for } |x| < \infty \quad (15b)$$

$$\partial_x G_2(x, z) \big|_{x=a} = 0 \quad \text{for } 0 < z < b \quad (15c)$$

$$\partial_x G_2(x, z) \big|_{x=-a} = 0 \quad \text{for } 0 < z < b \quad (15d)$$

Furthermore, G_1 and G_2 must satisfy the Sommerfeld radiation condition:

$$G_1, G_2(r, t) \sim e^{-i(\omega t - \vec{k} \cdot \vec{r})} \quad \text{as } |r| \rightarrow \infty \quad (16)$$

The equations of this section permit the complete evaluation of the three Green's functions. In regions I and III we find:

$$G_1(x, z; x', z') = \frac{1}{4} [H_0^{(1)}(k_1 R) + H_0^{(1)}(k_1 R')] \quad (17a)$$

$$G_2(x, z; x', z') = \frac{1}{4} [H_0^{(1)}(k_1 R) + H_0^{(1)}(k_1 R'')] \quad (17b)$$

where:

$$R = \sqrt{(x-x')^2 + (z-z')^2} \quad (18a)$$

$$R' = \sqrt{(x-x')^2 + (z+z'-2b)^2} \quad (18b)$$

$$R'' = \sqrt{(x-x')^2 + (z-z')^2} \quad (18c)$$

In region II, the method of images can be used. The result is given by Morse and Feshbach⁶:

$$G_2(x, z; x', z') = \frac{1}{4a\gamma_0} e^{i\gamma_0|z-z'|} + \frac{1}{2a} \sum_{m=1}^{\infty} \gamma_m^{-1} \cos\left[\frac{m\pi(x+a)}{2a}\right] \cos\left[\frac{m\pi(x'+a)}{2a}\right] e^{i\gamma_m|z-z'|}$$

where:

$$\gamma_m = \sqrt{k^2 - \left(\frac{m\pi}{2a}\right)^2} \quad \operatorname{Re}(\gamma_m) \geq 0 \quad \operatorname{Im}(\gamma_m) \geq 0 \quad (20)$$

With these explicit representations for G and the boundary conditions on both G and U , the integral representation for U given in Eq. (14) can be simplified. Furthermore, with the continuity conditions at the slit boundaries, U can be written exclusively in terms of U_2 and $\partial_z U_2$ at these boundaries. We find:

$$U^d(x, z) = - \int_{-a}^a \left(\frac{\epsilon_1}{\epsilon_2}\right) G_1(x, z; x', b) DU_b(x') dx' \quad \text{for } b < z < \infty \quad (21a)$$

$$U_1(x, z) = \int_{-a}^a \left(\frac{\epsilon_1}{\epsilon_2}\right) G_2(x, z; x', 0) DU_0(x') dx' \quad \text{for } -\infty < z < 0 \quad (21b)$$

$$U_2(x, z) = - \int_{-a}^a \left[G_2(x, z; x', 0) DU_0(x') - U_0(x') \partial_z G_2(x, z; x', z') \right]_{z=0^+} dx \\ + \int_{-a}^a \left[G_2(x, z; x', b) DU_b(x') - U_b(x') \partial_z G_2(x, z; x', z') \right]_{z=b^-} dx \\ \text{for } |x| < a \text{ and } 0 < z < b \quad (21c)$$

where the boundary fields are defined by:

$$U_0(x) \equiv U_2(x, z) \big|_{z=0^+} \quad (22a)$$

$$DU_0(x) \equiv \partial_z U_2(x, z) \big|_{z=0^+} \quad (22b)$$

$$U_b(x) \equiv U_2(x, z) \big|_{z=b^-} \quad (22c)$$

$$DU_b(x) \equiv \partial_z U_2(x, z) \big|_{z=b^-} \quad (22d)$$

Hence, the field can be found at any desired point once the four unknown functions in Eqs. (22) have been determined. The evaluation of these

functions is discussed below.

Development of Integral Equations for the Boundary Functions

A set of four integral equations is needed to completely determine the boundary functions. These equations result immediately from application of the continuity equations (11) and (12) to the field equations (21) at the slit boundaries. We find:

$$2U_b^1(x) - U_b(x) = \int_{-a}^a \left(\frac{\epsilon_1}{\epsilon_2}\right) G_1(x, b; x', b) DU_b(x') dx' \quad \text{for } |x| < a \quad (23a)$$

$$U_0(x) = \int_{-a}^a \left(\frac{\epsilon_1}{\epsilon_2}\right) G_2(x, 0; x', 0) DU_0(x') dx' \quad \text{for } |x| < a \quad (23b)$$

$$\begin{aligned} \frac{1}{2}U_b(x) = & \int_{-a}^a [G_2(x, b; x', 0) DU_0(x') - U_0(x') \partial_z G_2(x, b; x', z')|_{z=0+}] dx' \\ & + \int_{-a}^a G_2(x, b; x', b) DU_b(x') dx' \quad \text{for } |x| < a \quad (23c) \end{aligned}$$

$$\begin{aligned} \frac{1}{2}U_0(x) = & \int_{-a}^a [G_2(x, 0; x', b) DU_b(x') - U_b(x') \partial_z G_2(x, 0; x', z')|_{z=b-}] dx' \\ & - \int_{-a}^a G_2(x, 0; x', 0) DU_0(x') dx' \quad \text{for } |x| < a \quad (23d) \end{aligned}$$

where:

$$U_b^1(x) = e^{ik_1(x \sin \theta - b \cos \theta)} \quad |\theta| < \frac{\pi}{2} \quad (24)$$

These coupled equations for the four boundary functions will be solved numerically in the following section.

Approximate Solution of the Integral Equations

To demonstrate the technique that will be used to solve the system in (23), we consider Eq. (23a). The integral in this expression is first divided into N subintervals of width $2a/N$, each with a central point x_j given by:

$$x_j = \frac{2a(j - \frac{1}{2})}{N} - a \quad j = 1, 2, \dots, N \quad (25)$$

Furthermore, only the field at the midpoint x_k of each subinterval will be considered. Hence, we have:

$$2U_b^1(x_k) - U_b(x_k) = \sum_{j=1}^N \left(\frac{\epsilon_1}{\epsilon_2} \right) \int_{C_j} G_1(x_k, b; x', b) DU_b(x') dx' - \left(\frac{\epsilon_1}{\epsilon_2} \right) \int_{C_k} G_1(x_k, b; x', b) DU_b(x') dx' \quad (26)$$

The k^{th} integral is written independently since it contains the point x_k corresponding to the singularity in $G_1(x_k, b; x', b)$ as $x' \rightarrow x_k$.

Eq. (26) can be further reduced by applying the first mean value theorem to the $N-1$ subintervals C_j and the second mean value theorem to the subinterval C_k . This yields:

$$2U_b^1(x_k) - U_b(x_k) = \sum_{j=1}^N \frac{2a}{N} \frac{\epsilon_1}{\epsilon_2} G_1(x_k, b; \xi_j, b) DU_b(\xi_j) - \frac{\epsilon_1}{\epsilon_2} \left[\int_{C_k} G_1(x_k, b; x', b) dx' \right] DU_b(\xi_k) \quad (27)$$

The above expression is still exact for some unknown set of N points ξ_j , each within its own corresponding domain C_j . However, an approximation is now introduced to simplify the result. In particular, when the number of subintervals N becomes large, ξ_j can be estimated by the midpoint of each interval:

$$\xi_j = x_j \quad (28)$$

so that:

$$2U_b^1(x_k) - U_b(x_k) = \sum_{j=1}^N \frac{2a}{N} \frac{\epsilon_1}{\epsilon_2} G_1(x_k, b; x_j, b) DU_b(x_j) \quad k=1, 2, \dots, N \quad (29)$$

where:

$$s_{k,j}^I = \frac{2a}{N} \frac{\epsilon_1}{\epsilon_2} G_1(x_k, b; x_j, b) \quad \text{for } j=k$$

$$j, k=1, 2, \dots, N$$
(30a)

$$s_{k,k}^I = \frac{\epsilon_1}{\epsilon_2} \int_{Ck} G_1(x_k, b; x', b) dx' \quad (30b)$$

Eq. (29) can be written in matrix form:

$$2\hat{U}_b^I - \hat{U}_b = S^I D \hat{U}_b \quad (31)$$

where the components of the vectors are given by:

$$(U_b^I)_j = U_b^I(x_j) \quad j=1, 2, \dots, N \quad (32)$$

with similar expressions for $\hat{U}_b, D\hat{U}_b$ (and $\hat{U}_0, D\hat{U}_0$ later on). The matrix elements are found from the explicit representation of G_1 given in Eq. (17a):

$$s_{k,j}^I = i \frac{a}{N} \frac{\epsilon_1}{\epsilon_2} H_0^{(1)} \left(\frac{2k_1 a}{N} |k-j| \right) \quad \text{for } j \neq k \quad (33a)$$

$$s_{k,k}^I = i \frac{a}{N} \frac{\epsilon_1}{\epsilon_2} \left[H_0^{(1)} \left(\frac{k_1 a}{N} \right) + \frac{\pi}{2} \left\{ H_0 \left(\frac{k_1 a}{N} \right) H_1^{(1)} \left(\frac{k_1 a}{N} \right) - H_1 \left(\frac{k_1 a}{N} \right) H_0^{(1)} \left(\frac{k_1 a}{N} \right) \right\} \right] \quad (33b)$$

The Struve functions H_0 and H_1 result from an integration of the Hankel function $H_0^{(1)}$ about its singularity as outlined in Abramowitz and Stegun.⁷

Application of this same technique to the integral in Eq. (23b) yields a second matrix equation:

$$\hat{U}_0 = S^{III} D \hat{U}_0 \quad (34)$$

where:

$$s_{k,j}^{III} = i \frac{a}{N} \frac{\epsilon_1}{\epsilon_2} H_0^{(1)} \left(\frac{2k_1 a}{N} |k-j| \right) \quad \text{for } j \neq k \quad (35a)$$

$$s_{k,k}^{III} = i \frac{a}{N} \frac{\epsilon_1}{\epsilon_2} \left[H_0^{(1)} \left(\frac{k_1 a}{N} \right) + \frac{\pi}{2} \left\{ H_0 \left(\frac{k_1 a}{N} \right) H_1^{(1)} \left(\frac{k_1 a}{N} \right) - H_1 \left(\frac{k_1 a}{N} \right) H_0^{(1)} \left(\frac{k_1 a}{N} \right) \right\} \right] \quad (35b)$$

Note that S^{III} , like S^I , has only N independent elements and hence can be readily calculated.

Finally, the integrals in Eqs. (23c) and (23d) can be similarly approximated. After considerable calculation, we obtain:

$$\frac{1}{2}\hat{U}_b = -r^{II}\hat{D}\hat{U}_a + d^{II}\hat{U}_a + s^{II}\hat{D}\hat{U}_b \quad (36a)$$

$$\frac{1}{2}\hat{U}_a = -s^{II}\hat{D}\hat{U}_b + d^{II}\hat{U}_b + r^{II}\hat{D}\hat{U}_a \quad (36b)$$

where:

$$r_{k,j}^{II} = \frac{1}{2NY_0} e^{iY_0 b} + \frac{1}{N} \sum_{m=1}^N \frac{1}{Y_m} T_m^{j,k} e^{iY_m b} \quad (37a)$$

$$d_{k,j}^{II} = \frac{1}{2N} e^{iY_0 b} + \frac{1}{N} \sum_{m=1}^N T_m^{j,k} e^{iY_m b} \quad (37b)$$

$$s_{k,j}^{II} = \frac{1}{2NY_0} + \frac{1}{N} \sum_{m=1}^N \frac{1}{Y_m} T_m^{j,k} \quad (37c)$$

and

$$T_m^{j,k} = \frac{2N}{m\pi} \cos\left[\frac{m\pi(j-1/2)}{N}\right] \cos\left[\frac{m\pi(k-1/2)}{N}\right] \sin\left(\frac{m\pi}{2N}\right) \quad (37d)$$

In order to improve the convergence in Eqs. (37a)-(37c), the second mean value theorem was used on all N subintervals.

To summarize, the approximation in Eq. (28) has been used to transform the four coupled integral equations in Eqs. (23a)-(23d) into the four coupled matrix equations in Eqs. (31), (34), (36a), and (36b). The remaining goal is to solve this system for the vectors \hat{U}_a , \hat{U}_b , $\hat{D}\hat{U}_a$, and $\hat{D}\hat{U}_b$ which represent the values of the previously unknown boundary functions at N distinct points. Toward this end, \hat{U}_a and \hat{U}_b can be eliminated from the system to give a single $2N \times 2N$ matrix equation:

$$A \cdot \hat{B} = \hat{C} \quad (38)$$

where:

$$A_{mn} = \begin{cases} [0^{II} s^{III} - \kappa^{II}]_{m,n} & \text{for } m, n \in \{1, 2, \dots, N\} \\ [s^{II} + \frac{1}{2} s^I]_{m, n-N} & \text{for } m \in \{1, 2, \dots, N\} \quad n \in \{N+1, N+2, \dots, 2N\} \\ [s^{II} + \frac{1}{2} s^{III}]_{m-N, n} & \text{for } m \in \{N+1, N+2, \dots, 2N\} \quad n \in \{1, 2, \dots, N\} \\ [0^{II} s^I - \kappa^{II}]_{m-N, n-N} & \text{for } m, n \in \{N+1, N+2, \dots, 2N\} \end{cases} \quad \begin{matrix} (39a) \\ (39b) \\ (39c) \\ (39d) \end{matrix}$$

$$B_n = \begin{cases} (DU_0)_n & \text{for } n \in \{1, 2, \dots, N\} \\ (DU_b)_{n-N} & \text{for } n \in \{N+1, N+2, \dots, 2N\} \end{cases} \quad \begin{matrix} (39e) \\ (39f) \end{matrix}$$

$$C_n = \begin{cases} (U_b^I)_n & \text{for } n \in \{1, 2, \dots, N\} \\ [20^{II} \tilde{U}_b^I]_{n-N} & \text{for } n \in \{N+1, N+2, \dots, 2N\} \end{cases} \quad \begin{matrix} (39g) \\ (39h) \end{matrix}$$

Since A and \tilde{C} are composed completely of known elements, the unknowns \tilde{DU}_0 and \tilde{DU}_b can be readily determined. Substitution into Eqs. (36a) and (36b) then determines the remaining unknowns \tilde{U}_0 and \tilde{U}_b . In other words, by formulating the problem in the language of linear algebra, a discrete solution has been found for the complete diffraction problem. As the number of subdivisions N increases, this discrete solution should approach the true continuous solution.

There are several advantages to the method outlined here. First, the accuracy of the results can be increased without altering the basic algorithm by increasing the number of subintervals. Second, the validity of the solution is not limited to the narrow slit limit. Finally, the electromagnetic fields and energy flux can be determined using the simple

numerical approach described below.

Determination of the Near Field Energy Flux

When considering a slit as a light source for super-resolution microscopic applications, the degree of collimation of the radiation transmitted into region III is of great significance. Hence, the expressions below deal with quantities only in this region, although the same method can be used to find the fields in regions I and II as well.

The first mean value theorem can be applied to the integral expression for the field in Eq. (21b). If the approximation in Eq. (28) is then employed and the Green's function is written in explicit form, we find:

$$\vec{H}(x, z) = i \frac{a}{N} \frac{\epsilon_1}{\epsilon_2} \sum_{j=1}^N H_0^{(1)}(k, \sqrt{(x-x_j)^2 + z^2}) (DU_0)_j \hat{e}_y \quad (40a)$$

where the components $(DU_0)_j$ are found by determining \vec{B} in the matrix equation (38). From Maxwell's equations we then know:

$$E_x(x, z) = - \frac{a}{N} \frac{\sqrt{\epsilon_1}}{\epsilon_2} \sum_{j=1}^N \frac{z}{\sqrt{(x-x_j)^2 + z^2}} H_1^{(1)}(k, \sqrt{(x-x_j)^2 + z^2}) (DU_0)_j \quad (40b)$$

$$E_y(x, z) = 0 \quad (40c)$$

$$E_z(x, z) = \frac{a}{N} \frac{\sqrt{\epsilon_1}}{\epsilon_2} \sum_{j=1}^N \frac{x-x_j}{\sqrt{(x-x_j)^2 + z^2}} H_1^{(1)}(k, \sqrt{(x-x_j)^2 + z^2}) (DU_0)_j \quad (40d)$$

For field points very close to the screen, the second mean value theorem is needed to avoid the singularities in the Hankel functions.

Finally, it is desirable to present the radiation collimation data in the form of energy flux, since it is the intensity distribution that determines the resolution in near-field microscopy. Hence, the time averaged Poynting vector $\langle \vec{S} \rangle$ is the quantity of interest. For time harmonic fields, we find:

$$\langle S \rangle = \frac{c}{16\pi} [\vec{E} \times \vec{H}^* + \vec{E}^* \times \vec{H}] = \frac{c}{8\pi} \operatorname{Re} (\vec{E} \times \vec{H}^*) \quad (41)$$

Furthermore, since the object to be resolved will lie in a plane parallel to the screen containing the slit, we are most interested in the energy flux through such a plane, and thus concern ourselves primarily with $\langle S_z \rangle$.

Lastly, the flux should be normalized to the incident flux determined from Eq. (7). This yields:

$$\frac{S_z}{S_z^i} = - \frac{\sqrt{\epsilon_1}}{\cos \theta} \operatorname{Re}(E_x H^*) \quad (42)$$

Determination of Transmission Coefficients

Although the normalized transmission coefficient for a particular geometry could be calculated by integrating the above expression over the lower edge of the slit, an expression which is more tractable numerically can be obtained by manipulating the integral expressions for \vec{E} and \vec{H} . This is demonstrated below.

Integration of the normalized flux over the slit determines the transmission coefficient:

$$T = - \frac{\sqrt{\epsilon_1}}{4\cos \theta} \int_{-a}^a \lim_{z \rightarrow 0^-} [E_x H^* + E_x^* H] dx \quad (43)$$

Using the exact integral forms of E_x and H yields:

$$T = \frac{\epsilon_1 \epsilon_2}{4k_1 a \epsilon_2^2 \cos \theta} \lim_{z \rightarrow 0^-} [I_1 - I_2] \quad (44)$$

where:

$$I_1 = \int_{-a}^a dx \left[\int_{-a}^a \partial_z \partial_z^* (x, z; x'', 0) \partial_z \partial_z^* (x', 0; x'') dx'' \right] \cdot \left[\int_{-a}^a \partial_z \partial_z^* (x, z; x', 0) \partial_z \partial_z^* (x', 0; x'') dx' \right] \quad (45a)$$

$$I_2 = \int_{-a}^a dx \left[\int_{-a}^a \partial_z \partial_z^* (x, z; x'', 0) \partial_z^* \partial_z (x'', 0; x') dx'' \right] \cdot \left[\int_{-a}^a \partial_z \partial_z^* (x, z; x', 0) \partial_z (x', 0; x'') dx' \right] \quad (45b)$$

After several changes in the order of integration, these expressions reduce

to:

$$T = \frac{\epsilon_1 \epsilon_2}{4k_1 a \epsilon_2^2 \cos \theta} \int_{-a}^a dx'' DU_0(x') \int_{-a}^a dx' DU_0^*(x'') [I_+ - I_-] \quad (46)$$

where:

$$I_+(x', x'') = \lim_{z \rightarrow 0^-} \int_{-a}^a G_+(x, z; x'', 0) \partial_z G_+^*(x, z; x', 0) dx \quad (47a)$$

$$I_-(x', x'') = \lim_{z \rightarrow 0^-} \int_{-a}^a \partial_z G_-(x, z; x'', 0) G_-^*(x, z; x', 0) dx \quad (47b)$$

When the explicit form for G , is inserted, I_+ and I_- can be evaluated analytically. It is found that:

$$I_+ = \frac{1}{2} H_0^{(1)}(k_1 |x' - x''|) \quad (48a)$$

$$I_- = -\frac{1}{2} H_0^{(2)}(k_1 |x' - x''|) \quad (48b)$$

Substitution of these expressions into Eq. (46) and subsequent manipulation of DU_0 and DU_0^* leads to the final form of the transmission coefficient:

$$T = \frac{a^2 \epsilon_1 \epsilon_2}{N^2 k_1 a \epsilon_2^2 \cos \theta} \sum_{i=1}^N \sum_{j=1}^N J_0\left(\frac{2k_1 a}{N} |1-j|\right) [\text{Re}[(DU_0)_i] \text{Re}[(DU_0)_j] + \text{Im}[(DU_0)_i] \text{Im}[(DU_0)_j]] \quad (49)$$

where J_0 represents the zeroth order Bessel function. The elements $(DU_0)_j$ are obtained immediately when the system is solved, and J_0 is analytic, so that the transmission coefficient is easily obtained.

NUMERICAL METHODS

The accuracy and range of validity of this model is limited solely by the number of subdivisions N used in approximating the integral equations as a matrix equation. In particular, J_0 in the analog of Eq. (26) is roughly constant over the intervals Δ_j and hence the first mean value theorem can be applied only if the argument of the Bessel function in J_0 is roughly

constant. Hence, from Eqs. (18), we must have:

$$z > \frac{2a}{N} \quad (50)$$

for the approximations to be valid. Following the example of Neerhoff and Mur, $N=30$ was used to evaluate our boundary functions, so that the results could be applied to find the energy flux at distances greater than $1/30$ of the slit width beyond the screen. The resulting 60×60 matrix equations were solved using Gaussian elimination. The solution was checked for numerical accuracy by back substitution into the matrix equation, and was found to yield differences in the components of C of less than one part in 10^5 .

COMPARISON OF TRANSMISSION COEFFICIENTS

The first objective of this study was to check the consistency of the results by comparing the transmission coefficients determined from the integral approach of Neerhoff and Mur to those obtained by the method of Harrington and Auckland. Toward this end, Eq. (49) was used to calculate T for four different slit widths and a variety of screen thicknesses. The results are shown in Fig. 4.

From this figure, it is seen that the transmission resonances of $\lambda/2$ periodicity predicted by Harrington and Auckland are reproduced. Furthermore, the resonance positions are very similar, and the peaks heights are in agreement with the expression:

$$T = \frac{1}{ka} \text{ at resonance} \quad (51)$$

developed by Harrington and Auckland in the narrow slit limit. In short, the excellent agreement between these two disparate methods lends confidence in the validity of the results.

RESULTS AND IMPLICATIONS

Near-Field Scanning Optical Microscopy (NSOM)

To determine the degree of collimation within the near field, the boundary functions corresponding to the initial resonant thicknesses of Fig. 4 were used to calculate the energy flux in planes parallel to the screen. The results are shown in Figures 5a-5d for four different slit widths ($\lambda/2.5$ through $\lambda/20$). From these curves, the full width at half maximum (FWHM) of the energy flux perpendicular to the screen was determined as a function of the distance from the screen. These results are in Figure 6.

From the data in Figures 5 and 6, it is possible to estimate the feasibility of near-field microscopy at optical wavelengths. First, we note that the outward travelling radiation is collimated to exactly the slit width at the edge of the screen. This is a direct consequence of the perfect conductivity assumption, but is likely to hold in a realistic situation involving thin metal films of sufficient opacity. Hence, the basic concept of super-resolution microscopy remains valid: the light passing through an aperture in an opaque screen can be used to define a sub-wavelength light source which can then be used to generate an image by scanning.

Further analysis of Figures 5 and 6 indicates the practical difficulties inherent in near-field microscopy, due to the divergence of the radiation past the slit. As a rough criterion, the full width at half maximum of the radiation distribution is no more than 20% greater than the geometric projection of the slit at a distance of approximately half the slit width. In the visible regime ($\lambda \approx 5000\text{\AA}$), this implies that an aperture must be placed within 200\AA relative to a surface in order to obtain an image with $\approx 500\text{\AA}$ resolution. If we wish changes in the detected light intensity to be due to variations in optical properties of the sample, then the sharp decrease in the

peak flux as a function of distance places a further constraint on the precision of the aperture positioning. When scanning over low contrast samples such as biological specimens, this suggests that a positioning accuracy of $\approx 20\text{\AA}$ must be maintained in order to achieve $\approx 500\text{\AA}$ resolution using visible light.

These conclusions clearly illustrate the technical challenges inherent in extending the microwave results of Ash and Nicholls to the optical regime. However, due to recent developments in micropositioning technology such as seen in the scanning tunneling microscope⁸ and the capacitance microscope⁹, even the strict requirements of near-field optical microscopy can be met. Indeed, by employing micropositioning technology, we have already developed a prototype microscope, the Near-Field Scanning Optical Microscope (NSOM).^{10,11} With this instrument we have obtained better than 1000\AA resolution using a light source with $\lambda = 5000\text{\AA}$, thus circumventing the conventional diffraction limit. Fischer¹² and Pohl, et al.¹³ have also worked on the problem of near-field optical imaging, and have similarly concluded that subwavelength resolution is achievable. In addition Massey¹⁴ has investigated the infrared case, where the positioning requirements are less severe because of the larger wavelengths used.

Although half of the slit width can be used to provide an estimate of the collimation distance, it is also clear from Figure 6 that, even when the distance is scaled to the slit width, the radiation tends to diverge more rapidly as the width decreases particularly for the smaller widths. Hence, the apertures used for NSOM should not be made so small as to prohibit positioning of the sample within the collimated near-field. Limits on signal detection capabilities also provide a lower bound to the aperture size. On

the other, an upper bound is provided by the fact that the aperture should be at least as small as the desired resolution.

The existence of transmission resonances as a function of screen thickness leads to the question of what effect the screen thickness has on the spatial distribution of the energy flux. Presumably, the high transmission at resonance occurs when the system efficiently channels radiation from a wide area through the slit. In this case, one might assume that if the energy flow is symmetric about the screen, then the radiation should diverge very rapidly past the screen at resonance. Thus, the large signal strength associated with resonance could only be obtained at the expense of the resolution of the system. To test this hypothesis, energy flux distribution curves similar to Figures 5a-5d were developed for four slit widths corresponding to the anti-resonance positions in Figure 4. By comparing the two sets of data it was found that the resonance effect does not influence the collimation. To verify this unexpected result, the FWHM of the energy flux was calculated for the four different slit widths as a function of screen thickness at one particular distance from the screen. It was seen that, except for very small thicknesses, the distribution is indeed independent of thickness. Thus, high transmission can be achieved without a concurrent loss in resolution. In retrospect, this result is reasonable, since the symmetry of the problem is disrupted by the presence of the initial and reflected fields in addition to the diffracted field on one side (Eq. 9). As the thickness changes, the fields \vec{U}^d and \vec{U}^r change only in magnitude, but the field \vec{U}^t changes in distribution as well since it involves the sum of \vec{U}^d with the unchanging fields \vec{U}^i and \vec{U}^r . At resonance, the distribution of \vec{U}^t leads to a channeling of the radiation, but the distribution of \vec{U}^i remains unaffected.

Subwavelength Photolithography

The near-field collimation phenomenon can also be exploited to produce subwavelength structures via photolithography in the visible or near UV regime. However, because the collimation range is so short at optical frequencies, the pattern in a mask can be reproduced with high fidelity in photoresist only if the mask is in close proximity to the resist and only if the resist itself is very thin. Again referring to Figure 6, the mask should be roughly 200Å or less away from the furthest side of the resist in order to produce 500Å structures. This may be achievable by suitably thinning conventional resists, but should certainly be possible using recently developed Langmuir-Blodgett film technology.¹⁵

This method of super-resolution pattern generation has already been demonstrated by Fischer and Zingsheim.¹⁶ Using a conventional resist and a blue light source, they obtained 1000Å resolution, and with a photobleaching dye layer they obtained 500-700Å resolution. We are currently working to extend these results using sharper masks formed by electron beam lithography.

The model studied here is more applicable to lithography than to microscopy, since the masks used in the production of integrated circuits are generally formed from multiple slit-like structures, whereas the apertures used for two dimensional NSOM must be approximately round. However, the model is restrictive in that the magnetic field is assumed to be aligned along the slit. This leads to the existence of propagating modes within the slit (and correspondingly large transmission coefficients), but for other polarizations the modes may be evanescent. Thus, if polarized light is used for subwavelength photolithography, the exposure of the resist may depend on the orientation of the mask as well as on the sizes of the features incorporated therein. To avoid this problem, unpolarized illumination should be used.

The region between the mask and the substrate underlying the resist can also form a two dimensional waveguide which supports propagating modes.¹⁷ If monochromatic illumination is used, these modes can produce a series of $\lambda/2$ spaced fringes in the developed resist which are analogous to the transmission resonances shown in Figure 4. However, if polychromatic illumination is used, the field maxima under the mask should occur at different points, so that the fringes are eliminated by destructive interference. For example, a Xe arc lamp would be preferable to a laser or even to a Hg arc lamp because it has a smooth, broadband output in the visible and near UV.

High Density Optical Information Storage

The NSOM concept can be combined with the principle of subwavelength photolithography to produce high density read-write optical memories. To accomplish this, the photoresist must be replaced by a bistable optical material; that is, a material whose optical properties (e.g., absorption) can flip reversibly between two states based on the intensity and/or the frequency of the radiation employed. The light passed through an aperture is then used to change a small section of the surface to either the "on" or "off" state. The information in many such sections can then be read by using the same aperture with a different incident intensity and frequency as an NSOM probe. The bit size is limited only by the size of the aperture. Assuming that the bits are 500Å in diameter and are separated by 500Å, 10^{12} bits (more than 1 gigabyte) of information can be stored in 1 cm².

Near-Field Scanning Acoustic Microscopy (NSAM)

The above analysis of the near-field is also directly applicable to NSAM, Near-Field Scanning Acoustic Microscopy. In the acoustic case, the magnetic field U is replaced by the pressure P . This field still satisfies the Helmholtz equation if the excitation is harmonic:

$$(\nabla^2 - k^2) P = 0 \quad (52)$$

$$k = \sqrt{\rho} \frac{\omega}{c} \quad (53)$$

where ρ is the density of the fluid (analogous to the index of refraction). The velocity field \vec{u} is derivable from P just as the electric field \vec{E} is from the magnetic field, , except that it is longitudinal to the direction of travel rather than transverse to it:

$$\vec{u} = \frac{1}{\omega \rho} \nabla P \quad (54)$$

Furthermore, if the assumption of perfect conductivity is replaced by the assumption that the screen is perfectly rigid, then:

$$\partial_n P = 0 \quad \text{at the screen} \quad (55)$$

Comparing Eqs. (52) and (55) to Eqs. (3) and (9) it is seen that P satisfies exactly the same differential equation and boundary conditions as does U . Hence, the above numerical solution for the magnetic field immediately yields the pressure in the acoustic case, and the results concerning the collimation distance and the feasibility of super-resolution imaging remain valid.

NSAM has the potential to extend the resolution of conventional, diffraction limited acoustic microscopy^{1*} in the same manner that NSOM can extend optical microscopy. However, NSAM has two additional advantages. First, acoustic microscopy is limited in resolution by attenuation within the medium supporting the wave. Since one could obtain $\lambda/10$ resolution using the near-field technique, the frequency of the incident acoustic wave can be reduced by an order of magnitude with no concurrent loss of resolution. The dB attenuation, proportional to the frequency squared, can thus be reduced by a factor of 100 without loss of resolution if an NSAM system with a $\lambda/10$ aperture is employed. Second, a cylindrical acoustic waveguide, such as an

aperture in a thick rigid screen, supports at least one propagating mode no matter how narrow the constriction. Hence, attenuation of the signal within the aperture should be small compared to the optical case.

Topographical Mapping and Mask Alignment

The rapid decrease in the transmitted intensity past a slit can also be used to advantage in two ways. First, a non-destructive topographical instrument could be constructed by collecting light passed through a slit or aperture and reflected back through the aperture by an object within the near-field. As the aperture is scanned over an optically uniform surface, the intensity of the reflected light would represent changes in the aperture to object separation due to variations in the height of the object. Second, this same reflection scheme with a feedback mechanism could be used to position a lithographic mask with respect to a substrate.

The analysis of the near-field diffraction by a slit could be applied to a number of other problems as well ranging from electromagnetic interference between close packed semiconductor devices to the shielding of an environment from RF noise.

CAVEATS

The limitations of the above analysis must be considered before the results are used to apply near-field phenomena to experimental situations. First, it should be noted that the slit model was chosen more for its mathematical tractability than its correspondence to reality. For example, the metal films which can be used to form a thin screen have finite conductivities and hence are not completely opaque. This would increase the noise in an NSCM system due to the passage of a small fraction of the light in the region away from the aperture. Furthermore, the effective diameter of an

aperture would increase, since energy could propagate below the surface of the metal wall forming the aperture.

The model considered above is also limited to the situation where the incident magnetic field is aligned along the slit. For other polarizations, the transmitted intensity may be radically different, but the degree of collimation should be roughly the same.

In interpreting the results, it should be remembered that the full width of half maximum of the outgoing flux provides only a rough estimate of the resolution to be expected in a NSOM system. Of additional importance is the sharpness of the distribution below the slit, as is evident in Figures 5a-5d. Using this additional criterion, it is seen that the FWHM probably overstates the resolution, and that it may be necessary to place the aperture closer than half the width in order for the resolution to approach the aperture size.

Another potential problem concerns the effect of the object (for NSOM) or the resist and substrate (for lithography) on the distribution of radiation emanating from an aperture. Clearly, any such objects placed within the near-field change the boundary conditions and hence the solution of the entire problem. The most significant changes should occur for the case of conducting surfaces, but in all situations one might assume that the substrate should cause the radiation to spread more rapidly, since the screen and the surface can act as a waveguide to funnel radiation away from the aperture.

Although the principle of near-field imaging can be investigated using the infinite slit model studied here, a circular aperture must be used for two dimensional imaging in NSOM applications. In terms of the transmitted intensity, this can have significant consequences: the slit in a thick screen acts as a waveguide which can support a propagating mode, whereas a round aperture in a thick screen can support only evanescent modes if the aperture

is considerably smaller than a wavelength. Furthermore, there is an additional reduction in transmission due to reflections at the screen which occur even in the simpler case of an infinitely thin screen.

The geometry of an aperture closed in two dimensions has been considered by many researchers. Massey¹⁹ studied a square aperture in a thin screen using Fourier optics and considered the feasibility of super-resolution microscopy and lithography. The qualitative conclusions of the work concerning the extent of collimation are consistent with the more extensive distribution graphs in Figures 4a-4d, although the approach used may not be valid in the near-field. More recently, Leviatan²⁰ has determined the distribution of the magnetic field transmitted by a circular aperture in a thin screen using the method of moments. The collimation results were also found to be similar to the slit geometry studied here. The more realistic case of a round aperture in a thick screen has been considered by McDonald²¹, although only the degree of coupling between the two sides of the screen was investigated rather than the distribution of radiation in the near-field.

The results of McDonald can be used to obtain a rough estimate of the minimum aperture size which can be used for practical NSOM imaging. The least attenuated mode within a round aperture was found to be the TE_{11} mode, for which the energy decays at a rate of:

$$E = E_0 e^{-2 \cdot 1.841d/a} \quad (56)$$

where d is the screen thickness and a is the aperture radius. However, the energy decay in the screen at a finite rate as well, which is given by:

$$E = E_0 e^{-d/\delta} \quad (57)$$

where δ is the extinction length in the screen. The metal with the largest opacity in the visible is aluminum for which $\delta = 65\text{\AA}$ when $\lambda = 5000\text{\AA}$.²² When the

attenuation due to the waveguide effect exceeds the attenuation in the metal, the contrast between the aperture and the surrounding screen becomes insufficient for super-resolution applications. Hence, the minimum aperture diameter is given by:

$$2a = 4 \cdot 1.841 \cdot \lambda \quad (58)$$

For λ_1 , this yields a diameter of $\approx 480\text{\AA}$, which explains our choice of a goal of 500\AA for NSOM imaging. It should be noted that this is a very approximate result, since the finite conductivity of the screen should decrease the waveguide attenuation below the value given in Eq. (56) due to the larger effective aperture size as discussed above.

In a strict sense, the resolution of an NSOM system should indeed be limited to the aperture size, since two or more small structures will not be completely resolved if their separation is less than the aperture diameter. However, it should be possible to determine the position of the edge of a structure larger than the aperture (i.e., determine its position) to considerably less than the aperture size. To accomplish this, the intensity information obtained when the aperture is partially occluded by the edge would be deconvoluted to find the edge position.

Finally, although it is easiest to understand near-field microscopy in the scenario where the aperture is used to form a subwavelength sized light source, this is by no means the only possible mode for NSOM viewing. In addition, the light transmitted through the sample could be collected by the aperture in the near-field. In this arrangement, only the light from the region directly below the aperture is detected. Although the slit model suggests that this geometry may yield a lower resolution for a given slit to object separation due to the radiation scooping effect at resonance, the resolution should not be affected in the case of a circular aperture since all

modes are evanescent and no resonances occur. Finally, the light could be both transmitted by and collected with the aperture, thus forming a reflection NSOM instrument. This system has two principle advantages: even thick, opaque samples could be scanned, and the resolution and contrast may be higher than in the transmission modes, since the near-field collimation phenomenon is used to advantage not once but twice.

CONCLUSIONS

The energy flux transmitted by an infinite slit in a thick conducting screen has been determined using the method of Neerhoff and Mur. The model is limited only in the number of subdivisions taken to solve the resulting integral equations by numerical methods. Transmission coefficients were found to be identical to those calculated by the alternative method of Harrington and Auckland. The flux perpendicular to the screen is indeed collimated to the slit width within the very near-field, but diverges sufficiently rapidly so that an object must be placed closer than half the slit width in order to generate an image with a resolution approaching this width. In addition to near-field microscopy, other applications such as subwavelength photolithography and optical information storage were considered. Many physical phenomena which may be associated with the optical near-field were discussed qualitatively in conjunction with these applications. Finally, the limitations of the slit model were viewed in the context of its application to the design and construction of a practical Near-Field Scanning Optical Microscope.

ACKNOWLEDGEMENTS

The authors wish to thank Prof. Yehuda Leviatan for many profitable discussions. This work is supported by U.S. Air Force Contract #AFOSR-84-0314

and NSF Contract #ECS-8410304. One of the authors (EB) is supported as an IBM Graduate Fellow.

REFERENCES

1. H.A. Bethe, "Theory of Diffraction by Small Holes," Phys. Rev. 66, 163 (1944).
2. C.M. Butler, Y. Rahmat-Samii, and R. Mittra, "Electromagnetic Penetration Through Apertures in Conducting Surfaces," IEEE Trans. Antennas Propagat. AP-26, 82 (1978).
3. E.A. Ash and G. Nicholls, "Super-Resolution Aperture Scanning Microscope," Nature 237, 510 (1972).
4. R.F. Harrington and D.T. Auckland, "Electromagnetic Transmission Through Narrow Slots in Thick Conducting Screens," IEEE Trans. Antennas Propagat. AP-28, 616 (1980).
5. F.L. Neerhoff and G. Mur, "Diffraction by a Slit in a Thick Screen," Appl. Sci. Res. 28, 73 (1973).
6. P.M. Morse and H. Feshbach, Methods of Theoretical Physics (McGraw-Hill, New York, 1953) pp. 312-318.
7. M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions (Dover, New York, 1970) pp. 480, 496.
8. G. Binnig and H. Rohrer, "Scanning Tunneling Microscopy," Helvetica Physica Acta 55, 726 (1982).
9. J.R. Matey and J. Blanc, "Scanning Capacitance Microscopy," J. Appl. Phys. 57, 1437 (1985).
10. A. Lewis, M. Isaacs, A. Harootunian, and A. Muray, "Development of a 500Å Spatial Resolution Light Microscope," Ultramicroscopy 13, 227 (1984).

11. E. Betzig, A. Lewis, A. Harootunian, M. Isaacson, and E. Kratschmer, "Near-Field Scanning Optical Microscopy (NSOM): Development and Biophysical Applications," *Biophys. J.* 49 (1986, in press).
12. U.Ch. Fischer, "Optical Characteristics of 0.1 μ m Circular Apertures in a Metal Film as Light Sources for Scanning Ultramicroscopy," *J. Vac. Sci. Technol. B*, 3, 386 (1985).
13. D.W. Pohl, W. Denk, and M. Lanz, "Optical Stethoscopy: Image Recording With Resolution $\lambda/20$," *Appl. Phys. Lett.* 44, 652 (1984).
14. G.A. Massey, J.A. Davis, S.M. Katnik, and E. Omon, "Subwavelength Resolution Far-Infrared Microscopy," *Appl. Opt.* 24, 1498 (1985).
15. A. Barraud, "Polymerization in Langmuir-Blodgett Films and Resist Applications," *Thin Solid Films* 99, 317 (1983).
16. U.Ch. Fischer and H.P. Zingsheim, "Submicroscopic Pattern Replication With Visible Light," *J. Vac. Sci. Technol.* 19, 881 (1981).
17. K.L. Tai, R.G. Vadimsky, C.T. Kemmerer, J.S. Wagner, V.E. Lamberte, and A.G. Timko, "Submicron Optical Lithography Using an Inorganic Resist/Polymer Bilevel Scheme," *J. Vac. Sci. Technol.* 17(5), 1169 (1980).
18. B. Hadimioglu and C.F. Quate, "Water Acoustic Microscopy at Suboptical Wavelengths," *Appl. Phys. Lett.* 43, 1006 (1983).
19. G.A. Massey, "Microscopy and Pattern Generation With Scanned Evanescent Waves," *Appl. Opt.* 23, 658 (1984).
20. Y. Leviatan, private conversation.
21. N.A. McDonald, "Electric and Magnetic Coupling through Small Apertures in Shield Walls of Any Thickness," *IEEE Trans. Microwave Theory Tech.* MTT-20, 689 (1972).
22. J.W. Weaver, C. Kraska, D.W. Lyden, and E.E. Koch, "Optical Properties of Metals," *Physik Daten* 13, 1981.

FIGURE CAPTIONS

- Figure 1. Geometry of the Near-Field Scanning Optical Microscope.
- Figure 2. Model for diffraction by a slit in a thick screen.
- Figure 3. Regions of integration used for the boundary value problem.
- Figure 4. Transmission coefficients for $\lambda=5000\text{\AA}$ radiation as a function of screen thickness for slits of $\lambda/2.5$, $\lambda/5$, $\lambda/10$, and $\lambda/20$ widths.
- Figure 5. Distributions of S_z , the component of the energy flux normal to the screen, transmitted by slits of width: a) $\lambda/20$; b) $\lambda/10$; c) $\lambda/5$; d) $\lambda/2.5$ for $\lambda=5000\text{\AA}$ radiation.
- Figure 6. Rate of spreading of $\lambda=5000\text{\AA}$ radiation past the screen as determined by the full width at half maximum of S_z for slits of $\lambda/2.5$, $\lambda/5$, $\lambda/10$, and $\lambda/20$ widths.

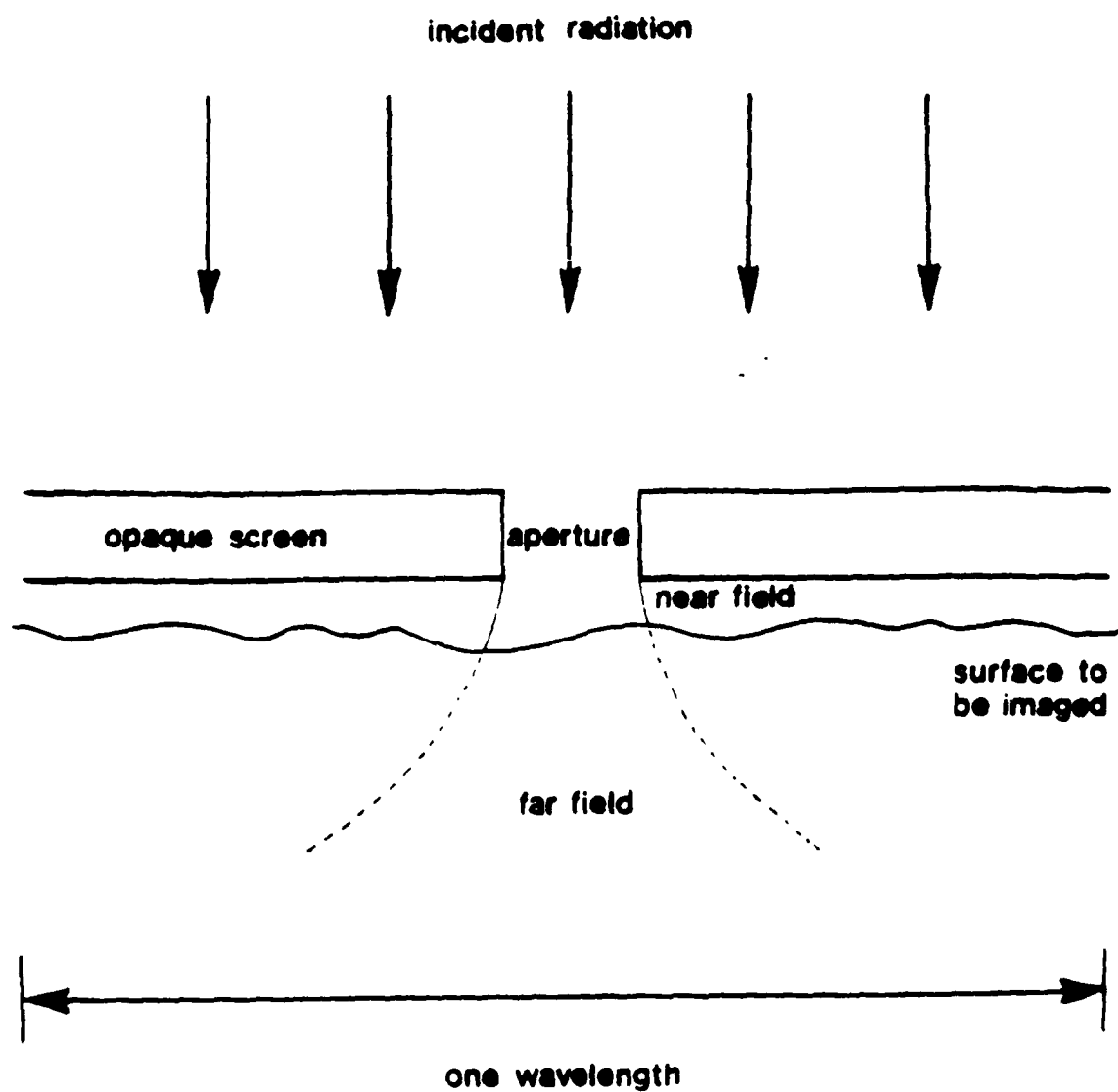


Figure 1:

Geometry of the Near-Field Scanning
Optical Microscope.

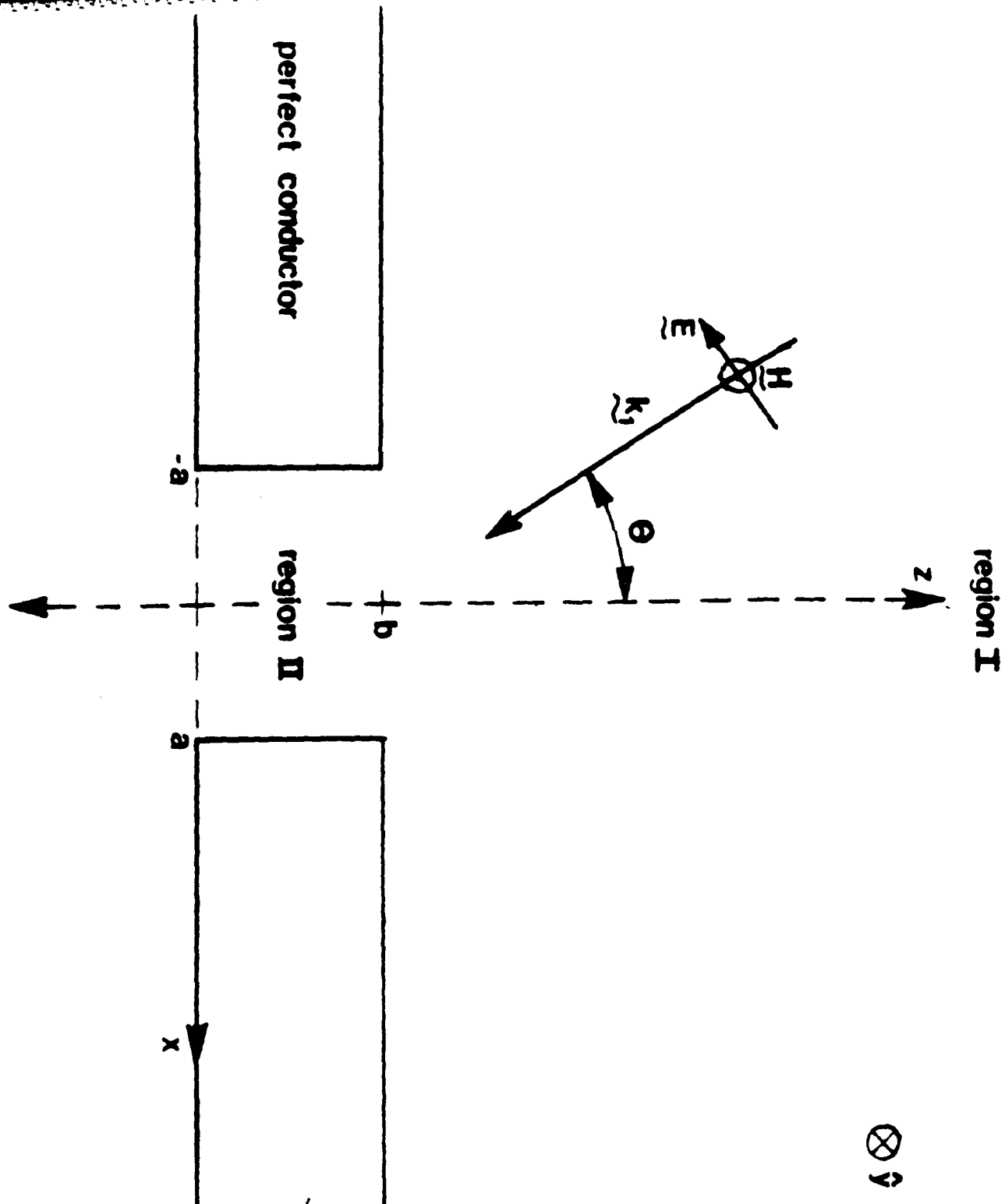


Figure 2:

Model for diffraction by a slit in a thick screen.

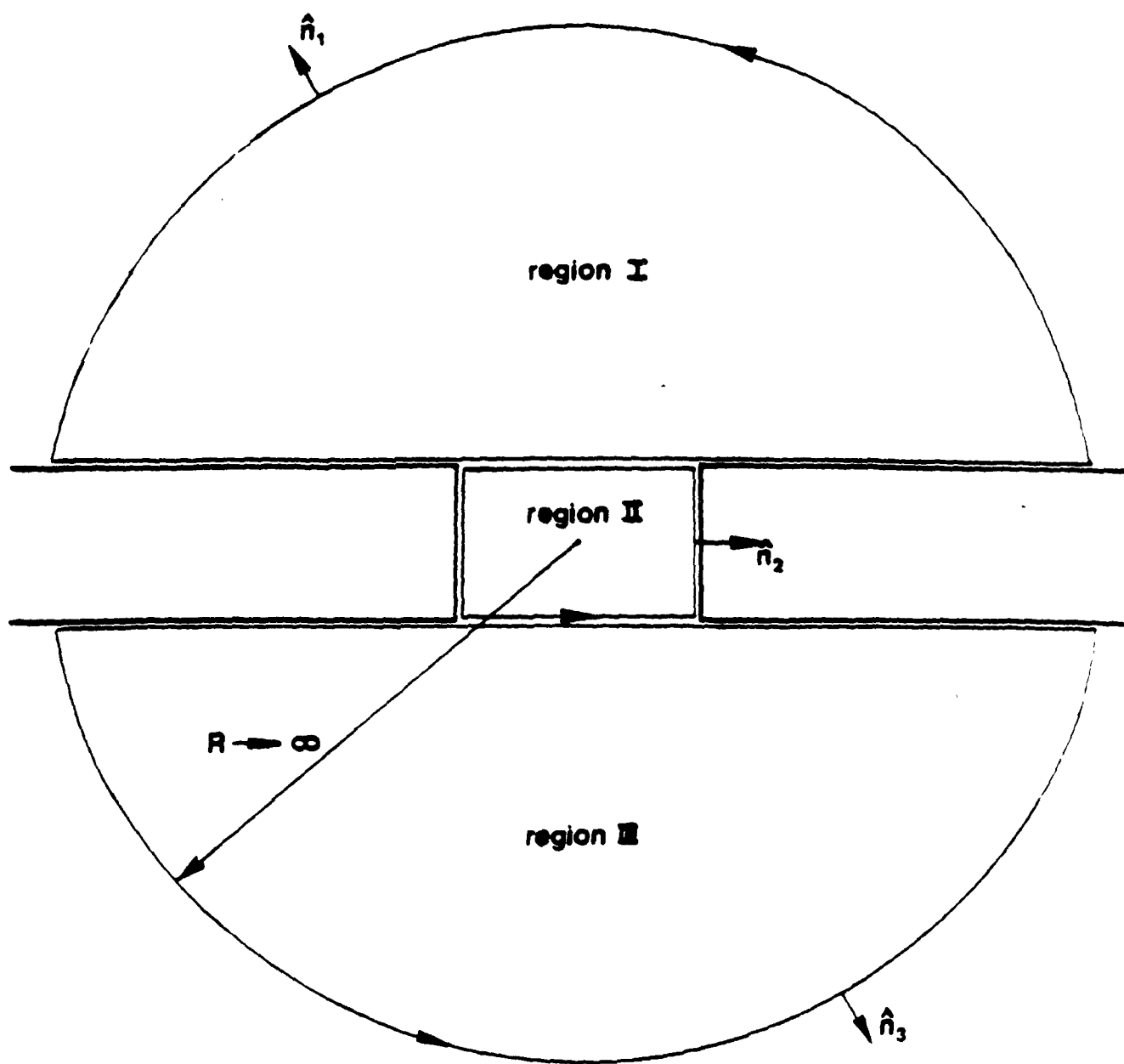


Figure 3:

Regions of integration used for the
boundary value problem.

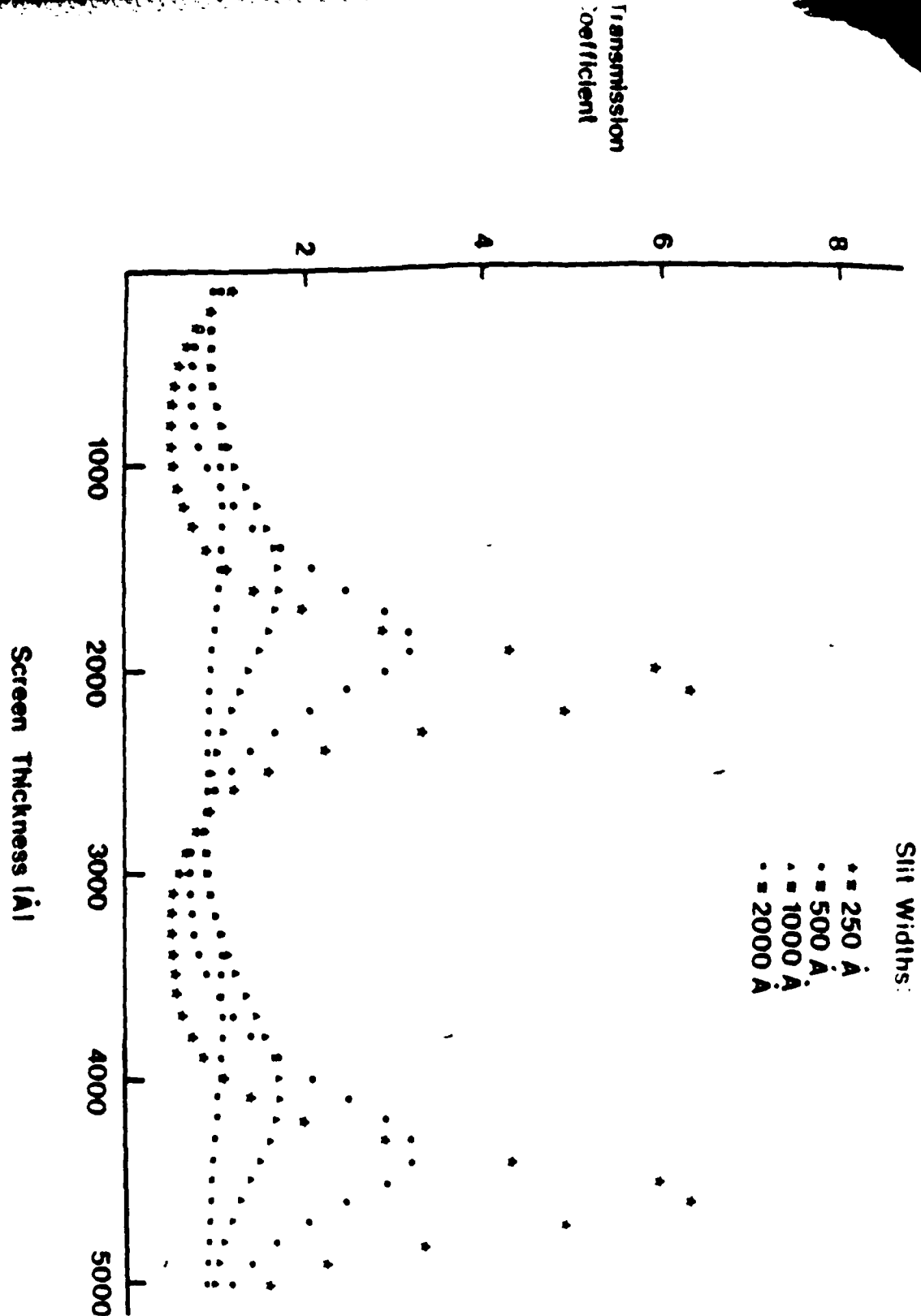


Figure 4:

Transmission coefficients for $\lambda=5000\text{\AA}$ radiation as a function of screen thickness for slits of $\lambda/2.5$, $\lambda/5$, $\lambda/10$, and $\lambda/20$ widths.

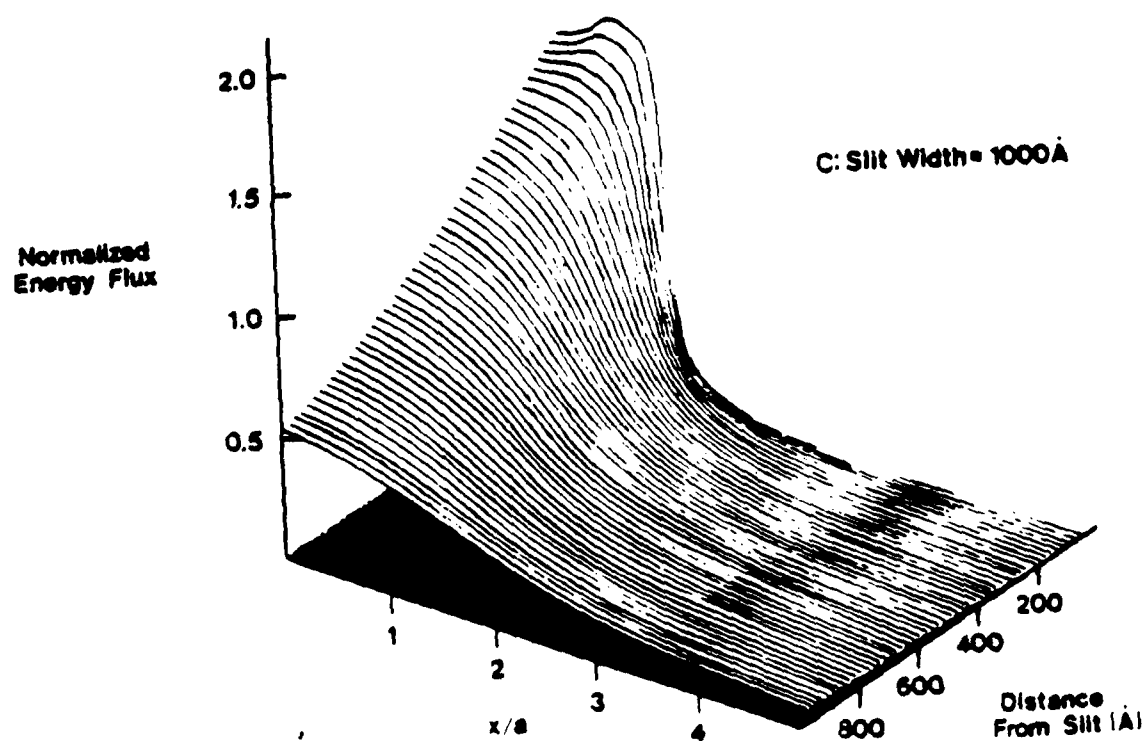
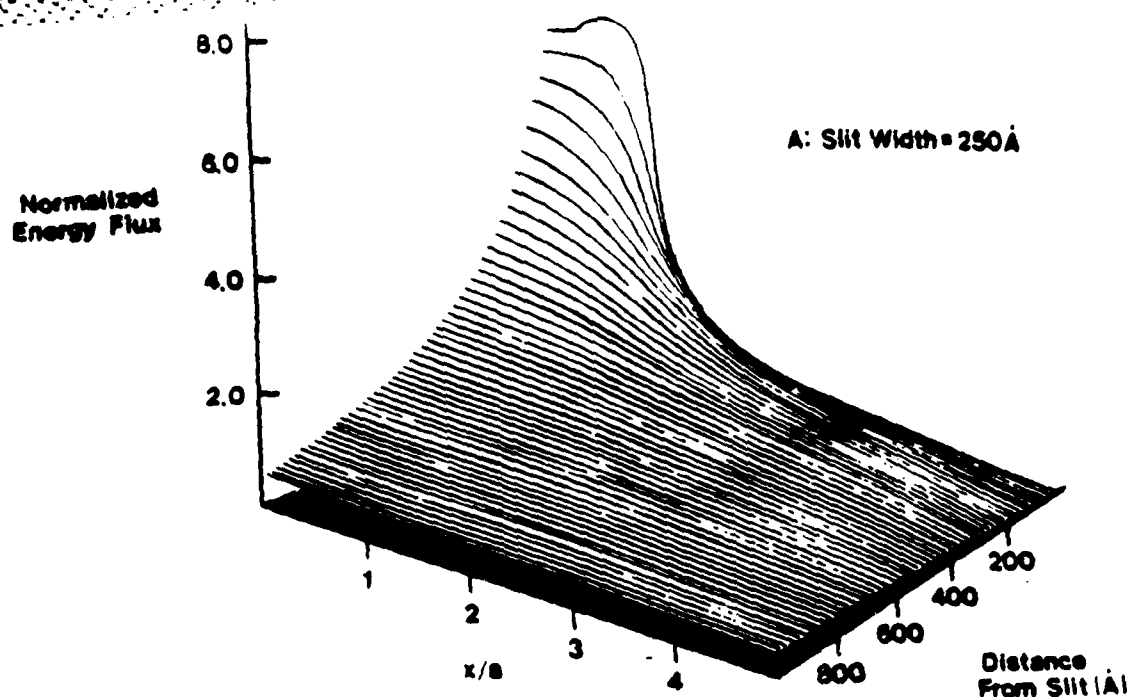


Figure 5:

Distributions of S_z , the component of the energy flux normal to the screen, transmitted by slits of width:
a) $\lambda/20$; b) $\lambda/10$; c) $\lambda/5$; d) $\lambda/2.5$
for $\lambda = 5000\text{\AA}$ radiation.

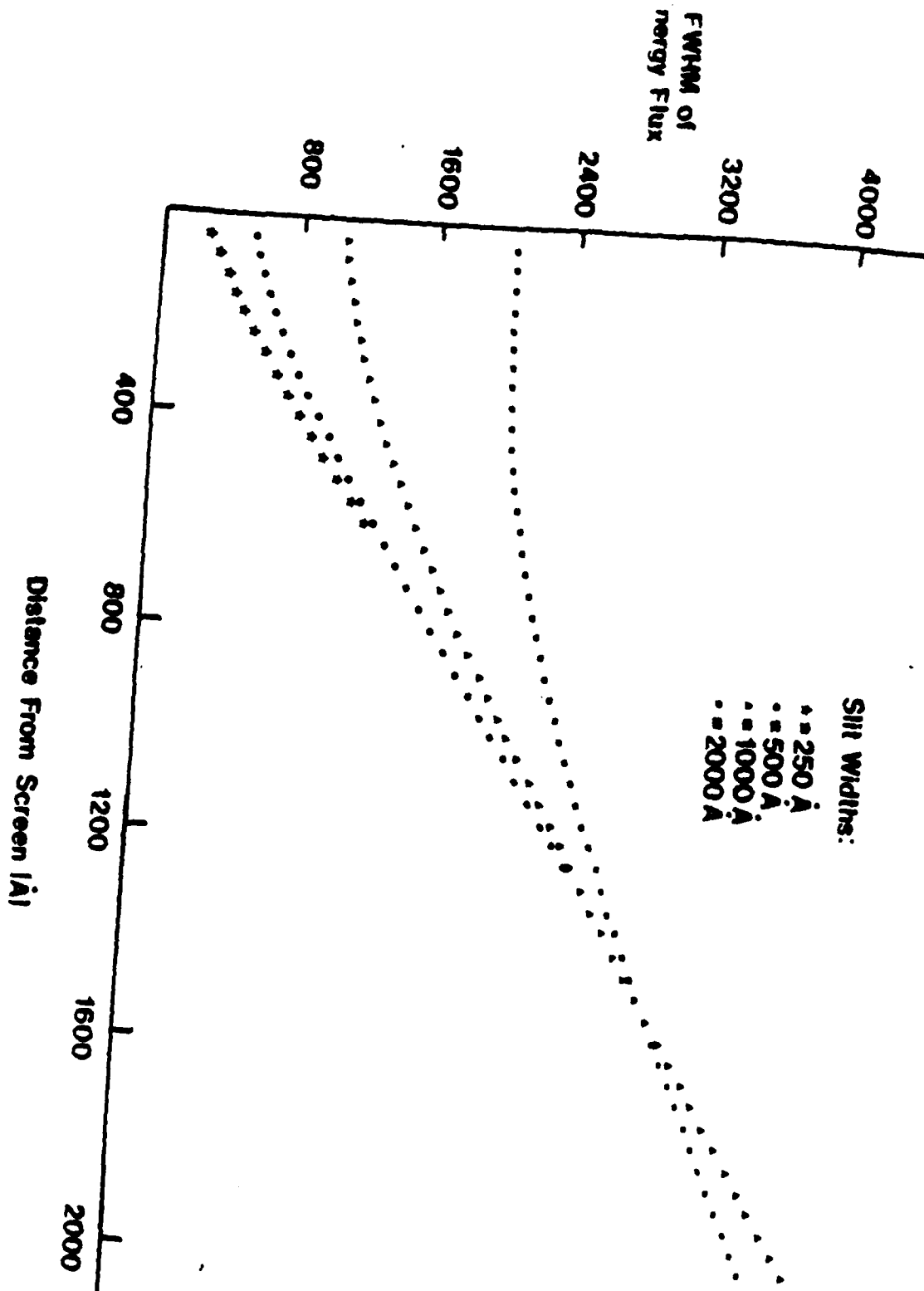


Figure 6:

Rate of spreading of $\lambda=5000\text{\AA}$ radiation past the screen as determined by the full width at half maximum of S_z for slits of $\lambda/2.5$, $\lambda/5$, $\lambda/10$, and $\lambda/20$ widths.

PART II

Near-Field Scanning Optical Microscopy (NSOM):

E. Betzig, A. Lewis, A. Harootunian, M. Isaacson, and E. Kratschmer

School of Applied and Engineering Physics

and

The National Research and Resource Facility for Submicron Structures

Cornell University
Ithaca, New York 14853

ABSTRACT

A new method for high resolution imaging, near field scanning optical microscopy (NSOM), has been developed. The concepts governing this method are discussed, and the technical challenges encountered in constructing a working NSOM instrument are described. Two distinct methods are presented for the fabrication of well characterized, highly reproducible, sub-wavelength apertures. A sample one dimensional scan is provided and compared to the scanning electron micrograph of a test pattern. From this comparison, a resolution of better than 1500\AA (i.e., $\approx \lambda/3.6$) is determined, which represents a significant step towards our eventual goal of 500\AA resolution. Fluorescence has been observed through apertures smaller than 600\AA and signal-to-noise calculations demonstrate that fluorescent imaging should be feasible. The application of such imaging is then discussed in reference to specific biological problems. The NSOM method employs non-ionizing visible radiation and can be used in air or aqueous environments for non-destructive visualization of functioning biological systems with a resolution comparable to that of scanning electron microscopy.

INTRODUCTION

Resolving components spatially in biological systems presents a classical biophysical problem. Most imaging techniques presently in use are fundamentally limited by the wavelength of the radiation. In this paper we present our development of a unique new imaging method that is independent of the radiation wavelength. This method permits the non-destructive imaging of surfaces within their native environment, with a resolution comparable to scanning electron microscopy. We have called this advance in biophysical technology, near field scanning optical microscopy (NSOM).

The resolution limits for this new technique fall between the extremes defined by light microscopy (2500Å) and fluorescence energy transfer techniques ($\approx 40\text{Å}$ to $\approx 80\text{Å}$). Using visible wavelengths, a conservative estimate of the resolution is approximately 500Å. Thus, it is likely that this technique will make an important contribution to the investigation of macromolecular assemblies such as lipid microdomains, receptor clusters, and photosynthetic reaction center aggregates. Furthermore, unlike other techniques at this resolution, near field imaging will be able to follow the temporal evolution of these macromolecular assemblies in living cells. Therefore, NSOM has the potential to provide both kinetic information and high spatial resolution.

The Conceptual Basis of Near Field Scanning Optical Microscopy (NSOM)

The fundamental principle underlying the NSOM concept is outlined in Figure 1, where visible light is depicted as being normally incident upon a conducting screen containing a small (sub-wavelength) aperture. Because the screen is completely opaque, the radiation emanating through the aperture and into the region beyond the screen is first collimated to the aperture size rather than to the wavelength of the radiation employed. This occurs in the near field regime. Eventually the effect of diffraction is evidenced as a

marked divergence in the radiation, resulting in a pattern that no longer reproduces the geometrical image of the aperture. This occurs in the far field regime.

To apply the collimation phenomenon, an object, such as a cell membrane, is placed within the near field region relative to an aperture. In this case, the aperture acts as a light source whose size is not limited by the considerations of geometrical optics. The light source can be scanned over the object, and the detected light can be used to generate a high resolution image. Because the resolution is dependent upon the aperture size rather than the wavelength, it should be possible to obtain 500Å or better resolution if a sufficiently small aperture is used. These are the essential features of the NSOM technique, which can be applied in air or aqueous environments using non-ionizing, visible radiation. Thus, macromolecular assemblies in functioning biological systems can be imaged at high resolution.

Comparison With Other Optical Methods

It is important to distinguish NSOM from more conventional diffraction limited techniques. For example, although images have been obtained of isolated 250Å structures using video enhanced contrast optical microscopy,¹ the images of these structures are, in the best of cases, limited by diffraction to sizes of 1000-2000Å. Hence, with this method it is difficult to determine the true sizes of the observed features. Furthermore, two or more structures will not be individually resolved if their respective Airy disks overlap (as determined by the Rayleigh criterion). In comparison, the NSOM method should generate images which indicate the true sizes of assemblies even when these assemblies are separated by much less than the wavelength.

Historical Background

As far back as a decade ago, the principle of super-resolution microscopy was demonstrated at microwave frequencies ($\lambda=3\text{cm}$) by Ash and Nicholls.² In

their pioneering experiment, a grating of 0.5mm periodicity was imaged with an effective resolution of $\lambda/60$. However, until we reported our initial results in 1983³, there were no published attempts to extend this technique to the visible region of the spectrum. This is quite understandable, because the minute physical dimensions of the optical near field demand aperture fabrication and micropositioning technologies on a nanometer scale. Furthermore, it is not immediately obvious that the results of the microwave experiment could be extended to the visible regime, because the finite conductivity of thin metal films has to be considered in the design of screens for submicron apertures, whereas the comparatively thicker metal screens used in the microwave case are completely conducting and thus opaque.

Since our original report, several additional publications on near field imaging technology have appeared.^{4 11} Fischer⁸ has produced an interesting series of results by scanning a sub-wavelength aperture over a second, larger aperture, but these results are difficult to interpret for a number of reasons. For example, the opacity of the metal films employed was not large, so that the apertures were poorly defined and much stray light was transmitted through the aperture screens. In addition, coherent, monochromatic illumination was used at grazing incidence, so that a series of standing waves may have been generated when the polarization of the electric field was perpendicular to these screens.

A near field imaging system for use in the far infrared is also currently being developed by Massey.¹⁰ Although this system may find many applications (such as in the detection of heat transport on a microscopic scale), it will not approach the resolution capabilities of NSOM and can best be viewed as a complementary imaging technique.

Finally, Pohl et al.¹¹ have developed a system for super-resolution microscopy that shares some of the same features as our NSOM instrument. They

call their technique optical stethoscopy. This is somewhat of a misnomer, as the analogy to a medical stethoscope is not strictly valid. This can be understood by treating the optical aperture and the acoustic stethoscope as classical waveguides. In the optical case, a consideration of the appropriate boundary conditions indicates that all modes are evanescent (decaying) within the aperture.¹² Indeed, only the finite thickness of the aperture screen permits the flow of any energy at all. A different set of boundary conditions must be invoked in the acoustic case. These conditions lead to the existence of a single propagating plane wave¹³ and explain why a 20Hz ($\lambda=17\text{m}$) sonic disturbance can be detected through a long stethoscope tube only a centimeter in diameter.

At the experimental level, the sizes and the structure of the apertures used by Pohl et al.¹¹ were not characterized. In addition, the nature of their manufacture presents considerable challenges in the attainment of reproducibility. To demonstrate 250Å resolution using optical techniques, not only must the apertures be well-characterized, but it is also essential to have well-characterized test structures. Neither of these criterion was met in Pohl et al.'s paper.

In short, initial attempts to implement the near field scanning concept attest to the difficulty of the NSOM technique. To further appreciate the technical challenges inherent in this form of microscopy, we have performed model calculations to obtain a first approximation to the pattern of radiation in the near field. These calculations (to be published) investigate the transmission of light through a slit of infinite length in a screen of finite thickness, and were based on the powerful Green's function technique of Neerhoff and Mur.¹⁴ In particular, the component of the time averaged Poynting vector perpendicular to the screen was calculated in the region just beyond the screen. The results are shown in Figure 2, where the divergence of

the energy flux is plotted as a function of the distance from the slit. These results demonstrate that the radiation remains collimated to a distance of at least half the slit width. Furthermore, they also indicate that the extent of the near field increases with the slit width. Finally, these near field energy flux calculations exhibit a close to exponential decrease in intensity with increasing distance from the screen. This suggests that rigid stability requirements will need to be maintained in the z direction, as will be discussed more fully in the following section. Of course, significant differences can be expected between the near field radiation pattern of a slit and that of a circular aperture, but the simpler slit geometry should give an indication of the pattern to be expected in the aperture case.

TECHNICAL CHALLENGES TO BE ADDRESSED

To construct a super-resolution microscope, several practical considerations must first be addressed. These considerations fall into two broad categories: micropositioning and enhancement of the signal to noise ratio.

Micropositioning

The theoretical analysis mentioned above indicates the degree of control required over the aperture to object separation. For example, the radiation emanating from a 500\AA wide slit remains collimated to an approximate distance of 250\AA . Furthermore, because the intensity in the near field is so strongly dependent upon distance, the separation must be maintained with a precision of about 20\AA . To satisfy such strict positioning requirements, several potential problems must be successfully confronted. These are discussed below.

Vibration Isolation

In a typical laboratory environment, building vibrations exist at frequencies as low as 2 to 4 Hz, although most of the vibrational energy is in

the 5 to 30 Hz range.¹⁵ In our environment, we have measured displacements of up to 1μ in this frequency range. If this energy were to be transferred without attenuation to the critical components of an NSOM system, and if the aperture and object were not held together with sufficient rigidity, then their relative displacement could vary by 2 to 3 orders of magnitude greater than the precision required. To preclude this possibility, it is necessary to design an isolation system which attenuates both horizontal and vertical vibrations from the floor at frequencies above a few Hertz. In addition, such a system must employ a damping mechanism to insure that any vibrations which are transmitted to the microscope are transient in nature. This damping mechanism is also needed to reduce the vibrations induced by the motion of the sample stage itself in the course of a scan. Finally, an effective isolation system must also shield the NSOM instrument from acoustic vibrations, which fall in the 20 Hz to 20 kHz range.

Thermal Drift

Although the thermal expansion coefficients of the various components of an NSOM system will vary widely with composition and size, it can be expected that these components will expand/contract roughly 0.1 to 1.0 microns for every 1°C increase/decrease in temperature. Hence, the differential rate of expansion of the aperture relative to an object could present a serious obstacle towards obtaining the required micropositioning capabilities for super-resolution microscopy. This obstacle can be surmounted through two lines of attack: either the instrument can be designed so as to insure that the thermal expansion of the aperture relative to the object is rather small, or else stringent control can be maintained over the temperature of the entire apparatus. Ideally, both of these methods should be used.

Aperture and Sample Translation

As a final micropositioning requirement, it is necessary to translate the aperture and sample accurately relative to one another in all three spatial dimensions. As mentioned above, the translating system should position the aperture with a precision of roughly 20\AA in the z direction (perpendicular to the sample plane). However, it is equally necessary to obtain accurate positioning in the x and y directions, as one factor which limits the resolution of an NSOM system is the size of the steps taken while scanning. In particular, the step size should be shorter than half of the desired resolution.¹⁶ Hence, a system designed for 500\AA resolution must include positioning control of better than 250\AA in the x and y directions.

Factors Determining the Signal to Noise Ratio

Signal Considerations

The light transmitted through a submicron aperture is weak enough to require the use of sensitive detection electronics. A good detection system becomes increasingly important when it is considered that, in the interest of producing high resolution images rapidly, the aperture must scan quickly over the object, so that the time available at each point to accumulate the requisite counting statistics is rather brief. Weak signals can particularly be expected in biological applications, where contrast differences in a specimen can be quite small. In such applications, a variety of non-destructive methods would be needed to increase contrast, ranging from the computer enhancement of images to the fluorescent labeling of specimens.

Noise Considerations

Since rather weak signals are expected, one obvious source of noise is that of statistical noise. The significance of this source can be reduced by

increasing the period of data collection, but this is done at the expense of scan speed. Systematic noise can also be expected from the detection system in the form of dark noise and noise from the preamplifier and discriminator electronics. Noise will also be introduced if there is any variation in the light intensity used to illuminate the sample or aperture. Furthermore, since the aperture screen is not completely opaque, any light which is transmitted through this screen (rather than through the aperture) will contribute to the systematic noise.

There are also sources of noise related to the issue of micropositioning. First, unwanted variations in the xy positioning of the sample relative to the aperture can either cause an apparent distortion in the image (for variations long compared to the scan time) or else can smear the contrast and resolution information at a point (for variations short compared to the data collection time at that point). Of even more significance are the large variations in apparent signal strength which can be caused by slight changes in the aperture to object separation. Finally, noise can be introduced when scanning thick ($>1000\text{\AA}$), translucent samples, because the system will detect light scattered from the diffraction limited far field regions as well as from the collimated near field regime. In short, considerable effort must be expended to increase signal strength and reduce noise so as to permit rapid high resolution microscopy.

INITIAL FEASIBILITY STUDIES

Our initial goal was to construct a simple prototype to demonstrate the feasibility of the NSOM concept. The essential features of this prototype are discussed below.

Submicron Aperture Fabrication

Aperture Arrays

Two radically different techniques have been employed in fabricating submicron apertures for near field microscopic studies. The first scheme consists of the formation of aperture arrays in a thin, planar silicon nitride membrane. These arrays were formed using electron beam lithography at the National Research and Resource Facility for Submicron Structures. This technique, which was described in an earlier paper¹⁷, has generated uniform, reproducible holes as small as 80Å in diameter. After aperture formation, the opacity of the surrounding membrane was increased by evaporating aluminum over the arrays. Similar apertures have already been applied in optical transmission studies in which we have demonstrated that visible light from a tungsten source is transmitted through holes as small as 300Å.³ The efficiency of the transmission process is demonstrated in Figure 3A, where light transmission is shown through 2400Å diameter apertures. Figure 3B demonstrates the alteration in the signal to noise ratio in going to a 600Å diameter aperture array. The uniformity of the transmitted intensity attests to reproducibility of this method of aperture fabrication. In addition to the utility of such arrays in characterizing the optical properties of submicron apertures, they can be applied to non-scanning near field imaging methods.

Pipette Apertures

The principal problem in near field imaging is the finite depth of field due to the limited extent of the collimation of radiation in the near field. Because of this limitation, apertures in a planar membrane cannot probe recessed regions in rough surfaces. To address this problem, we have developed a simple, inexpensive method of aperture fabrication involving the formation of submicron apertures in the tips of highly tapered, metallized glass pipettes.

This method is based on an extension of the technology developed for making glass pipettes used in electrically probing cells by the patch clamp method.¹⁸ We have extended this technology by: a) fabricating apertures with diameters of less than 1000\AA , b) characterizing the apertures at the tips of the pipettes with scanning electron microscopy, c) demonstrating that light is transmitted through such pipette apertures.

A two stage method was used for constructing the apertures. The first stage tapered the pipette and the second stage pulled until the breakage point was reached. By varying the wall thickness of the glass pipettes, aperture diameters from less than 1000\AA to 5000\AA with outer diameters of 5000\AA to 7500\AA were reproducibly generated. Thicker walled pipettes were found to produce smaller apertures. Aluminum was again evaporated to increase the opacity of the glass wall surrounding the aperture. This evaporation resulted in a uniformly metallized tapered pipette with an aperture smaller than 1000\AA at the tip (see Figure 4).

The taper of the pipette is of particular importance. When light is passed down the pipette, it is transmitted through the outer glass walls as well as through the central region (see Figure 5). The inner diameter of the pipette rapidly tapers to dimensions of less than an optical wavelength. This model of light transmission suggests that the radiation exists in a propagating mode throughout the length of the pipette, since the pipette outer diameter is larger than the cut-off value calculated by treating it as a classical wave guide. Only at the thin metallized region at the tip of the pipette is the cut-off threshold reached. Thus, the large throughputs observed are a consequence of the fact that the region of evanescence is short. Upon leaving the aperture, the radiation once again exists in a propagating mode.

An NSOM Prototype--Design

The pipette type apertures were an important advance in our development of a simple NSOM prototype, but even for the initial feasibility experiments, the technical challenges discussed above had to be addressed in at least a preliminary fashion. The consideration of these challenges in our development of a workable NSOM prototype will now be discussed.

Initial System

A two-stage system was employed to shield the instrument from floor vibrations. The first consisted of a large table top mounted on tennis balls. Measurements in our laboratory demonstrated that this system reduced high frequency vertical disturbances to amplitudes of less than 300Å. The second consisted of a heavy steel plate resting on a foam pad, which provided isolation in the intermediate frequency range (10 to 50 Hz).

Using precision translation stages and conventional micrometers, the pipette was positioned with approximately one micron accuracy in the xy plane over the section of the sample to be scanned. Another stage was used to position the pipette within a few microns of the sample surface. To bring the pipette within the near field region, a stacked piezoelectric element was used (see Figure 6). A low-noise (15mV peak to peak), high-voltage supply was employed to insure that the error in the piezoelectric expansion due to voltage ripple was considerably less than the 20Å limit mentioned previously. Another piezoelectric element was used to drive the pipette in the scan direction. The step size could be varied, but was typically 150Å. Because conventional piezoelectric elements have limited expansion ranges, the scans were constrained to be no larger than several microns. Furthermore, these conventional elements exhibit hysteresis problems ($\approx 15\%$ of the full

expansion), so that accurate positioning could be achieved only along a single axis.

To deal with the problems posed by signal to noise considerations, the prototype system (see Figure 6) employed an optical multichannel analyzer with an SIT vidicon detection system in conjunction with a conventional light microscope. This design insured that a sufficient portion of the available light was collected and detected. Noise in each channel of the detector was at most half the signal due to the aperture, even when small apertures and weak (tungsten) illumination were used. Much of the noise consisted of a constant background level, which was easily subtracted. To measure the noise from other sources, the final detection signal was plotted as a function of time in a test situation where the aperture position was fixed. The resulting variations in intensity were roughly one percent of the apparent signal. These noise fluctuations were at least an order of magnitude smaller than the intensity changes observed in the course of a super-resolution scan.

NSOM Prototype--Results

Microfabricated Patterns for Resolution Tests

To test the NSOM method, it was necessary to fabricate objects that could be used to accurately define the resolution limits of the resulting device. Toward this end, electron beam lithography was used to create a mask consisting of several sets of line gratings with varying periodicity. The image of this mask was reproduced with high fidelity as a series of aluminum lines on a glass coverslip using the method of contact printing. These lines were 2000Å wide and were separated by spacings ranging from 2000Å-8000Å. In Figure 7, a scanning electron micrograph of one of these gratings is shown. Note the quality with which the pattern is reproduced. As noted above such

high-quality test objects with reproducible features are required before any conclusions can be drawn about NSOM resolution.

A One Dimensional Scan

Using the NSOM prototype and the test object discussed above, a one dimensional scan was performed which demonstrated super-resolution capabilities. The light from a 100W tungsten lamp was passed through a pipette to illuminate the aperture. The tip was brought into contact with the sample surface and then retracted several nanometers for scanning.

The initial scan was performed using a 5000Å diameter pipette type aperture and the results are shown in Figure 8. Note that the periodicity of the grating pictured in Figure 7 is reproduced throughout the entire length of the NSOM scan. A resolution of better than 1500Å is inferred from averaging the sharpness of the steps in going from transmission minima to adjacent transmission maxima and vice versa. Note further that the technique permits detection of sharp edges with a resolution considerably better than the limit expected based on a naive model of the aperture size. This is understandable, because with a sufficient signal to noise ratio it should be possible to detect changes in intensity due to partial occlusion of the aperture by the sharp edge.

Future Improvements

Although the prototype system was sufficient to demonstrate the super-resolution concept, it suffered from a number of limitations. These included the inability to scan in two dimensions, the inability to scan over rough topography, the inability to produce highly reproducible scans, and the inability to scan at high speed. Methods exist for attacking all these problems. We are currently incorporating these methods in the construction

of a second generation NSOM instrument. Many of the ideas employed in this construction effort were first developed for scanning tunneling microscopy, where positioning on a sub-Ångstrom scale is required.¹⁹ At the completion of this effort, we hope to have a versatile system capable of investigating problems of biophysical interest.

BIOPHYSICAL APPLICATION

Fluorescence Detection

Because NSOM employs visible radiation, many of the techniques which have already been developed for conventional microscopy (e.g., polarized illumination, phase contrast methods, etc.) can be used to advance the biophysical applicability of super-resolution microscopy. In particular, contrast enhancement by fluorescence techniques will be useful, because many biological systems can be selectively labeled with fluorophores in a non-destructive fashion or else contain fluorescent domains in their natural state. The only potential problem associated with fluorescent labeling lies in the reduction of the signal strength due to the limited cross sections and quantum yields of these fluorophores. However, in this paper, we have taken a first step in demonstrating the feasibility of fluorescence imaging by observing fluorescence emanating from sub-wavelength apertures (see figure 9). This result was obtained by illuminating a thin film of perylene with 5145Å radiation from a laser, and viewing the fluorescence transmitted through an aperture array with the appropriate filter. Note that relatively strong signals are obtained, even though perylene has only a moderate quantum efficiency.

Fluorescence From a Functioning Biological System

The next step in approaching a significant biological problem involved observing the transmission of fluorescent light through submicron apertures

emanating from a viable biological specimen. To achieve this, spinach chloroplasts were used. A suspension of such chloroplasts was placed directly on an array of 2400Å holes, and the suspension was made sufficiently dense so that the chloroplasts adhered to the array and did not migrate. The chloroplasts were illuminated through the apertures using 4579Å laser radiation incident from the opposite side of the array, and the fluorescence was detected back through these same apertures in reflection. The results, shown in Figure 10, demonstrate the feasibility of detecting such fluorescence on a submicron scale in a non-destructive manner.

Chloroplasts were chosen for our initial experiments because they are involved in the current debate concerning the distribution of the two photosystems involved in photosynthesis. Photosystem II is involved with the splitting of water while photosystem I is involved with the alteration in the oxidation-reduction potential of electrons ejected from the split water. Both photosystems are found in the thylakoid membranes of chloroplasts, but only photosystem II fluoresces at room temperature.

The membranes themselves contain two structurally distinct regions: the grana, in which the membranes are stacked; and the stroma, which are regions of unstacked membranes interconnecting the grana regimes. According to one model, photosystem I is excluded from the stacked regions of the grana. However, the evidence on which this model is based comes from freeze-fracture electron microscopy, detergent methods of fractionating the grana and stroma membranes, and mechanical fractionating methods^{20,21}, all of which have a disruptive effect on the membranes. Conventional microscopy cannot be used, since the grana regions cannot be resolved with sufficient clarity to determine the distribution of the two photosystems. In contrast, near field microscopy should provide the resolution necessary to determine the photosystem distribution within viable chloroplasts maintained in their native

environment. We are currently working on this problem, the results of which should help elucidate energy transfer mechanisms involved in photosynthesis.

Signal to Noise Ratio Calculations for Biological Applications

The signals detected by fluorescence NSOM imaging are weak and many possible sources of noise can be expected. In this section we demonstrate the feasibility of fluorescence imaging methods by calculating the signal to noise ratio in a specific biological application. The particular problem we focus on is the detection of a receptor in the plasma membrane of human fibroblast cells that binds low density lipoprotein (LDL). LDL is the predominant transporter of plasma cholesterol, and at least 30 fluorescent lipid molecules can be incorporated per LDL molecule.²² It has been shown that with the fluorescent lipid 3, 3'-dioctadecylindocarbocyanine iodide (diI[3]) single LDL molecules on single LDL receptors can be detected.²² We can use this data to obtain signal to noise estimates which will indicate the feasibility of imaging such receptors. However, these calculations are not limited to this example, because phycobiliproteins have been shown to be as highly fluorescent as LDL and can be used to selectively label a variety of cellular receptors.²³ We have been able to irradiate LDL molecules with green light for about one minute at an intensity of $\mu\text{W}/\mu^2$ before the resulting fluorescent image begins to fade. This corresponds to a total dose of 1.9×10^{22} photons/cm².

To determine how this limit affects the incident intensity and hence the total signal in an NSOM application, we consider here the specific case of near field microscopy in reflection. The use of such a geometry is advantageous, because each molecule is affected only when the illuminated aperture is overhead. In practice, this is not strictly true, because the finite opacity of the screen must be taken into account. For example, a 600Å thick layer of aluminum will transmit roughly 0.01% of the incident light. If

the attenuation of the radiation in the aperture is assumed to be 90%, then the differential transmittability of the aperture relative to the screen is 10^3 .

In a typical scan, we may wish to record a 100×100 element image within 100 seconds. In this case, the differential transmittability indicates that the LDL molecules will receive dosages roughly ten times greater from transmission through the screen as they will from transmission through the aperture. Hence, to remain below the photobleaching threshold within the 100 second scan time, a maximum photon flux past the screen of 1.7×10^{20} photon/sec/cm² is permissible, corresponding to an incident intensity of $5.4 \text{ mW}/\mu^2$.

Although at first glance it may appear troubling that the dosage from screen transmission is an order of magnitude greater than that from aperture transmission, when the situation is considered in detail it is not problematical at all. First, it must be remembered that these figures refer to the total dosages in the course of a scan, and at any given time the intensity directly below the aperture is roughly 1000 times greater than the intensity below the screen. Furthermore, from the standpoint of the signal to noise ratio, we are only interested in the fluorescent light which is transmitted back through the aperture and screen for eventual detection. In as much as this is the case, the transmission of light through the screen is unimportant, because any fluorescence which might otherwise be generated by LDL molecules under the screen is highly quenched by the presence of the metal within the screen.²⁴ Hence, only the fluorescence from the near field directly below the aperture contributes to the signal, and only far field fluorescence contributes to the noise.

The maximum permissible intensity calculated above can now be used to estimate the total signal which can be expected in our Gedanken experiment.

In particular, by again assuming a 10% transmittability for the aperture and noting that the above scan parameters lead to a 10ms exposure time per pixel location, we find that 1.7×10^{21} photons/cm² for each pixel are transmitted through the aperture. We will further assume that these photons are all incident on the fluorescent site of a single receptor.

To calculate the resulting signal, we note that each of the 30 diI chromophores per LDL molecule has an effective photoabsorption cross section of 3.8×10^{-17} cm² with a fluorescent yield of 0.1.²² Using these numbers, we find an initial signal of 1.9×10^5 fluorescent photons/pixel.

A number of factors cause the detected signal to be lower than this initial value. First, in the reflection mode described here, the assumed aperture transmission reduces the fluorescent light intensity by 90%. Second, inefficiencies in the collection optics can be assumed to further reduce the intensity by 50%. Finally, the photomultiplier that will form the heart of a sensitive detection system has a quantum efficiency of only 20%. When all these factors are taken into account, the detected signal is found to be 1.9×10^3 counts/pixel.

To determine the noise associated with a signal of this strength, we first consider noise due to photon counting. Statistical noise for the above counting level is roughly 44 counts/pixel. In contrast, dark noise from the photomultiplier should be less than 10 counts/sec. For the above scan rates, this corresponds to 0.1 counts/pixel, and hence is quite negligible.

Other fundamental sources of noise arise from the limited extent of the collimation of radiation in the near field. For example, although the LDL molecules in close proximity to the metal screen will not exhibit fluorescence, other such molecules bound to the far side of the cell membrane will be weakly illuminated by residual far field radiation from the aperture, and hence may yield dim fluorescence noise. This noise is considerably less

than the signal due to fluorescence from receptors within the intense near field radiation pattern emanating from the aperture, as can be understood by the following simple argument.

As was mentioned previously, the intensity of radiation emanating from a slit falls in an almost exponential manner within the near field. It is reasonable to assume that a similar behavior exists in the case of a circular aperture. In fact, by treating an aperture of diameter $d=500\text{\AA}$ as a waveguide, a specific form of $\exp[-2\pi r/d]$ can be assumed to hold out to a collimation distance of roughly 250\AA . Hence, the radiation should remain collimated through the entire thickness of that portion of the cell membrane which is directly under the aperture, and any fluorescence that occurs in this region can be considered to contribute to the signal. In contrast, any fluorescence which occurs in the far field region contributes to the noise. However, within the far field, the aperture can be modeled as a dipole radiator, so that the outward flux falls as $1/r^6$. If we assume that this behavior is valid up to $r=250\text{\AA}$, then the continuity of intensity at this distance can be used to provide a complete, qualitative model of the energy flow past the aperture.

We can now use such a model to estimate the signal-to-noise ratio due to the limited extent of the collimated near field. The signal is determined by integrating the exponential intensity factor over a hemisphere of 250\AA radius. Similarly, by integrating the factor of $1/r^6$ over a 250\AA thick shell one micron away (which represents the membrane furthest from the aperture), the noise is found. In this example, which should fairly well model the binding of LDL receptors to fibroblast cells, the noise is found to be several orders of magnitude less than the signal. Even if we were to assume that chromophores are evenly distributed throughout the bulk of the cells (which is certainly not the case in our example), a signal to noise ratio of 7.7 can be obtained using this model. When it is also considered that the fluorescence

from the chromophores decreases in intensity by $1/r^6$ as well, the noise is dropped even further.

The implications of this last point cannot be overemphasized. The finite extent of the near field is used to advantage since the far field noise is small compared to the near field signal. Thus, NSOM can be used to image surfaces on thick samples without sectioning.

Returning to our model for imaging LDL receptors, it can now be seen that the most significant source of noise is statistical in nature. Thus, by using the above results, we find that a signal-to-noise ratio of about 44 should be obtainable, and hence the experiment should be entirely feasible. Indeed, it would not be necessary to run the experiment at the threshold for photobleaching. If, for example, the incident intensity were to be decreased by a factor of ten, a signal to noise ratio of almost 14 would still be achieved. Alternatively, the size of the scan speed could be increased by a factor of ten with an identical change in the signal-to-noise ratio.

CONCLUSIONS

Our initial prototype system has demonstrated the feasibility of NSOM as a high resolution imaging technique. Furthermore, the results of the fluorescence experiments and the signal-to-noise ratio calculations indicate that NSOM may be applied to a variety of biological problems. For example, with the development of a flexible second generation system, NSOM could be used to: image lipid microdomains in model membrane assemblies²⁵ or living cells; investigate banding patterns in chromosomes with fluorescent molecules that bind to particular DNA sequences and structures²⁶; delineate cytoskeletal interactions with membrane bound proteins; characterize alterations in cell surface topography as a result of cellular stimulation²⁷; visualize organelle movement along dissociated actin filaments²⁸; and to study

the kinetics of specific proteins during muscle contraction.²⁹ In addition, this technique could be used to investigate problems in many other areas from microelectronics to chemistry where non-destructive, high resolution imaging is required.

ACKNOWLEDGEMENTS

The chloroplast imaging experiments are in collaboration with Prof. Chanoch Carmeli. We would like to thank Brian Whitehead for electron beam exposure of the mask gratings.

This work is supported by U.S. Air Force Contract #84-NE-121, U.S. Army Contract #DAMD17-79-C-9041, and the National Research and Resource Facility for Submicron Structures through NSF Contract #EC58200312.

FIGURE CAPTIONS

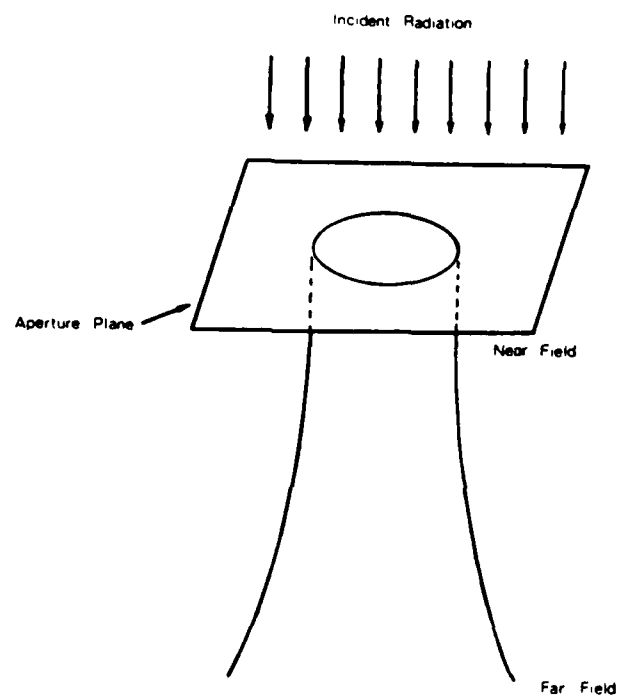
- Figure 1 Schematic representation showing the collimation of radiation emanating from a sub-wavelength aperture.
- Figure 2 The energy flux emanating from a 500\AA wide slit, normalized to the energy of the incident plane wave, is shown plotted as a function of both the perpendicular distance (z) from an 1800\AA thick conducting screen and the ratio of the lateral distance from the center of the slit to the half width of the slit (x/a). Notice that the radiation is collimated to roughly the edge of the slit ($x/a=1$) out to approximately $z=250\text{\AA}$.
- Figure 3 (A) 5145\AA light transmission through 2400\AA diameter apertures is shown detected in the far field with an optical multichannel analyzer. Note the uniformity of the light transmission through the nine apertures shown. Incident beam power was 0.62 m W/cm^2 . (B) 5145\AA light transmission through 600\AA diameter apertures is shown as detected in the far field with an optical multichannel analyzer. Incident laser power was as above. Note the even for these smaller apertures a good signal-to-noise ratio is achieved.
- Figure 4 A scanning electron micrograph of the tip of a metallized pipette is shown with a 5000\AA outer diameter and with a $<1000\text{\AA}$ central aperture. Although even smaller apertures could be prepared with our technique, they would be difficult to characterize with the limited resolution of the SEM available.
- Figure 5 A cross-sectional representation of a metallized pipette is shown together with the propagation of light toward the tip.

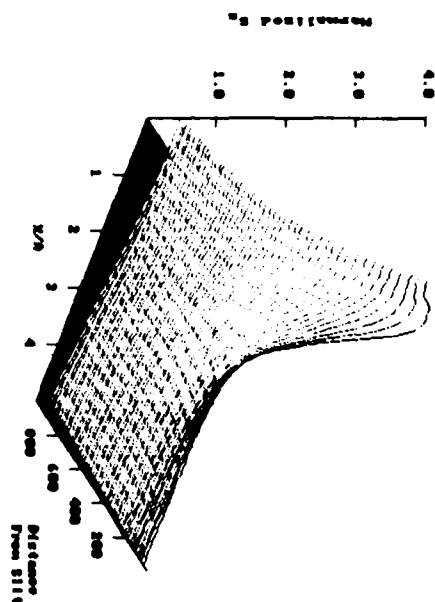
- Figure 6 A schematic diagram of the experimental apparatus used for our initial feasibility experiments is shown. Both the optical multichannel analyzer and the piezoelectric translators are controlled with a microcomputer.
- Figure 7 A scanning electron micrograph of a grating of aluminum lines on glass used for our resolution tests is shown. The lines were 2000Å separated by 4000Å. To obtain this micrograph the sample was coated with a 100Å thick film of gold palladium.
- Figure 8 A scan of the test grating in Figure 7 is shown. From the sharpness of the intensity modulation between adjacent maxima and minima, a resolution of <1500Å is inferred. Furthermore, note that the periodicity of the grating is reproduced over the entire scan.
- Figure 9 Perylene fluorescence through (a) 1200Å diameter and (b) 600Å diameter apertures are shown. The exciting laser light at 5145Å illuminated the sample from below, and the fluorescence was detected through the apertures from above. The laser power that was used caused no detectable bleaching of the sample. The fluorescence was recorded with ASA 400 Kodak Ektachrome film using a 30 sec exposure. These arrays contain graduated aperture sizes with the maximum corresponding to the diameters noted above.
- Figure 10 Photosystem II fluorescence is shown detected through 2400Å diameter apertures using epi-illumination. The incident intensity from an argon ion laser at 4579Å was $0.067 \mu\text{W}/\mu^2$ at the surface of the array. Kodacolor ASA 400 film and an exposure time of 40 sec was used to detect the fluorescence from the chloroplasts.

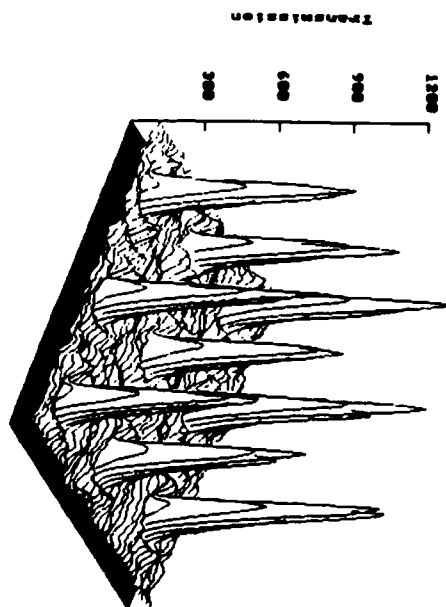
REFERENCES

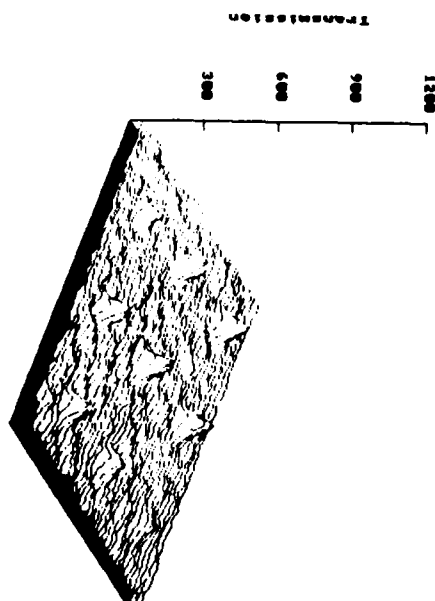
1. R.D. Allen, New observations on cell architecture and dynamics of video-enhanced contrast optical microscopy, *Ann. Rev. Biophys. Chem.* 14, 265 (1985).
2. E.A. Ash and G. Nicholls, Super-resolution aperture scanning microscope, *Nature* 237, 510 (1972).
3. A. Lewis, M. Isaacson, A. Muray and A. Harootunian, Scanning optical microscopy with 500Å spatial resolution, *Biophys. J.* 41, 405a (1983).
4. A. Lewis, M. Isaacson, A. Harootunian and A. Muray, Development of a 500Å spatial resolution light microscope, *Ultramicroscopy* 13, 227 (1984).
5. A. Harootunian, E. Betzig, A. Muray, A. Lewis, and M. Isaacson, Near field investigation of submicrometer apertures at optical wavelengths, *J. Opt. Soc. of Amer. A* 1, 1293 (1984).
6. E. Betzig, A. Harootunian, A. Lewis and M. Isaacson, Near field scanning optical microscopy (NSOM), *Biophys. J.* 47, 407a (1985).
7. E. Betzig, A. Harootunian, E. Kratschmer, A. Lewis and M. Isaacson, Non-destructive optical imaging of surfaces with 500Å resolution, *Bull. Amer. Phys. Soc.* 30, 482 (1985).
8. U. Ch. Fischer, Optical characteristics of 0.1µm circular apertures in a metal film as light sources for scanning ultramicroscopy, *J. Vac. Sci. Technol. B* 3, 386 (1985).
9. G.A. Massey, Microscopy and pattern generation with scanned evanescent waves, *Applied Optics* 23, 658 (1984).
10. G.A. Massey, J.A. Davis, S.M. Katnik and E. Omon, Subwavelength resolution far-infrared microscopy, *Appl. Opt.* 29, 1498 (1985).
11. D.W. Pohl, W. Denk and M. Lanz, Optical stethescope: image recording with resolution $\lambda/20$, *Appl. Phys. Lett.* 44, 652 (1984).
12. J.D. Jackson, *Classical Electrodynamics*, 2nd ed., John Wiley and Sons, New York, NY, p. 344 (1975).
13. L.E. Kinsler, A.R. Frey, A.B. Coppens and J.V. Sanders, *Fundamentals of Acoustics*, 3rd ed., John Wiley and Sons, New York, NY, p. 218 (1982).
14. F.L. Neerhoff and G. Mur, Diffraction by a slit in a thick screen, *Applied Scientific Research* 28, 73 (1973).
15. Newport Research Corporation Catalog, Fountain Valley, CA, 13 (1983-84).
16. C.E. Shannon, Communication in the presence of noise, *Proc. IRE* 37, 10 (1949).

17. A. Muray, M. Isaacson, I. Adesida and B. Whitehead, Fabrication of apertures slots and groves at the 8-80nm scale in silicon and metal films, J. Vac. Sci. Tech. B1, 1091 (1983), M. Isaacson, A. Muray, M. Scheinfein, I. Adesida, and E. Kratschmer, Nanometer structure fabrication using electron beam lithography in Proceedings of the International Symposium on Nanometer Structure Electronics, Osaka, April 1984, eds. K. Gamo, Ohmsha, Ltd., Tokyo (1985 in press).
18. Single Channel Recording eds. B. Sakmann and E. Neher, Plenum Press, New York City, NY (1983).
19. G. Binnig and H. Rohner, Scanning tunneling microscopy, Helvetica Physica Acta 55, 726 (1982).
20. J.M. Anderson, Consequences of spatial separation of photosystem 1 & 2 in thylakoid membranes of higher plant chloroplasts, FEBS Lett. 124, 1 (1981).
21. J. Barber, An explanation for the relationship between salt-induced thylakoid stacking and the chlorophyll fluorescence changes associated with changes in spillover of energy from photosystem II to photosystem I, FEBS Lett. 118, 1 (1980).
22. L.S. Barak and W.W. Webb, Fluorescent low density lipoprotein for observation of dynamics of individual receptor complexes on cultured human fibroblasts, J. Cell Biology 90, 595 (1981).
23. V.T. Oi, A.N. Glazer and L. Stryer, Fluorescent phycobiliprotein conjugates for analyses of cells and molecules, J. Cell Biology 93, 981 (1982).
24. K.H. Drexhage, Interaction of light with monomolecular dye layers, Progress in Optics 12, 165 (1974).
25. S.W. Hui and D.F. Parsons, Direct-observation of domains in wet lipid bilayers, Science 190, 383-384 (1975).
26. J.K. Barton, L.A. Basiles, A. Banishefsky and A. Alexandrescu, Chiral probes for the handedness of DNA helixes: enantiomers of tris(4, 7-diphenylpheanthroline)ruthenium (II). Proc. Natl. Acad. Sci. U.S.A. 81, 7 (1984).
27. J.M. Fernandez, E. Neher and B.D. Gomperts, Capacitance measurements reveal stepwise fusion events in degranulating mast cells, Nature 312, 453 (1984).
28. B. Kachar, Direct visualization of organelle movement along actin filaments dissociated from Characean algae, Science 227, 1355 (1985).
29. A.P. Somlyo and A.V. Somlyo, Smooth muscle structure and function in Handbook of Experimental Cardiology, eds. H.M. Fozzard, E. Haber, R.B. Jennings, A.M. Katz, and H.E. Morgan, Raven Press, New York (in press).

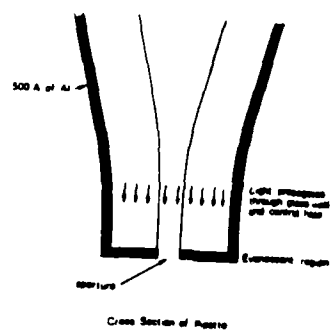


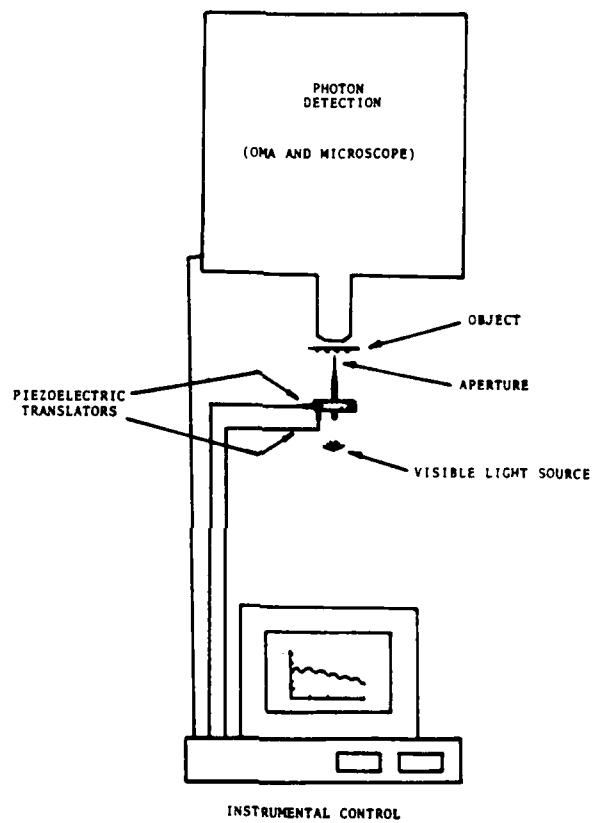


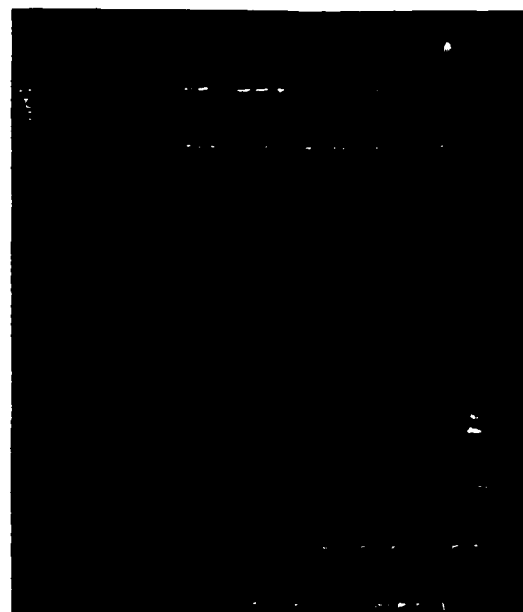


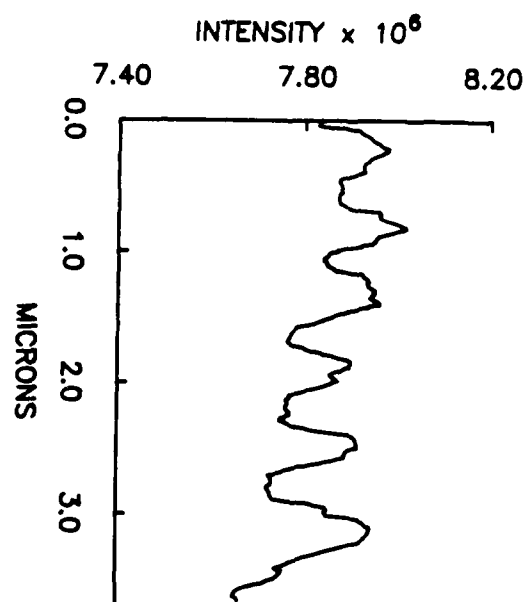


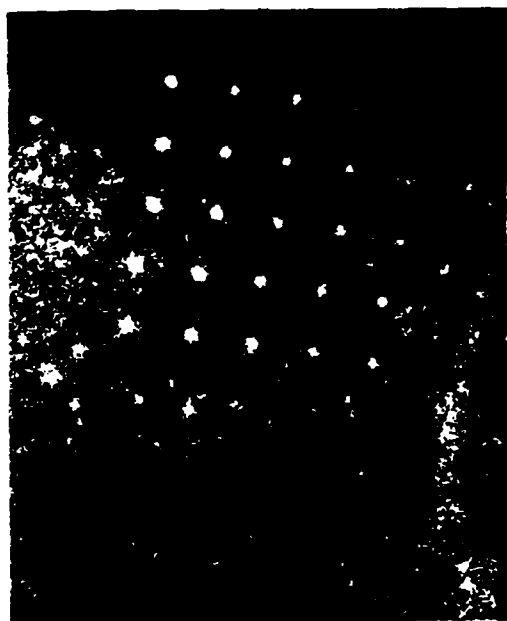




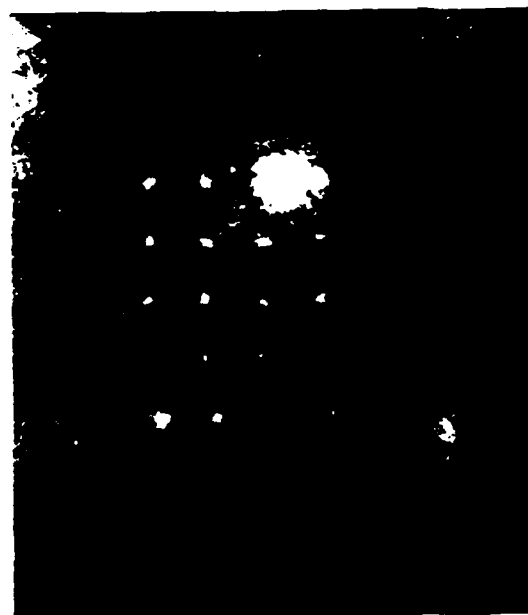












END

10-86

DTIC